

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Kant on Freedom, Nature and Normativity

Permalink

<https://escholarship.org/uc/item/4kd4m30m>

Author

Kohl, Markus

Publication Date

2012

Peer reviewed|Thesis/dissertation

Kant on Freedom, Nature and Normativity

By

Markus Kohl

A dissertation submitted in partial satisfaction of the

Requirements of the degree of

Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Hannah Ginsborg, Co-Chair

Professor Jay Wallace, Co-Chair

Professor Kinch Hoekstra

Spring 2012

© 2012
Copyright
Markus Kohl
All rights reserved

Abstract

Kant on Freedom, Nature and Normativity

By

Markus Kohl

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professor Hannah Ginsborg, Co-Chair

Professor Jay Wallace, Co-Chair

In typical debates about freedom, compatibilists and incompatibilists dispute whether and how one can defuse the threat that a naturalistic worldview seems to pose for our conception of free will. Proponents of both sides standardly take it for granted that natural causality is a potential threat only for *practical* deliberation, and that no corresponding worries arise for our self-conception as *theorizers* who make up their mind about what to believe about the world.

I show that Kant rejects this standard dichotomy between practice and theory. He holds that we must both *act* and *believe* 'under the idea of freedom': the standpoint of theoretical reasoning, the *very source* of the naturalistic worldview, must itself rely on the idea of freedom (or spontaneity) and thus cannot expose free will as an illusion. This line of thought, I argue, casts a refreshing new light on debates about what kind of freedom is worth having; and it provides an intriguing defense of our freedom from natural necessity.

I argue that Kant's conception of theoretical freedom is as anti-naturalistic as his conception of practical freedom. This twofold anti-naturalism derives from Kant's conviction that both our practical and our epistemic orientation depend on representations that we cannot attribute to natural causes without losing our grip on their objective validity. These include, in the practical case, the categorical imperative and, in the theoretical case, the principle of sufficient reason, which states that natural events are causally necessitated. Since we do not derive these representations from sense perception, our commitment to them must, within a wholly naturalistic framework, be traced to the influence of contingent psychological processes in the vein of Hume's 'customs'. The concession that our attitudes are the product of such customs challenges our conviction that we form these attitudes on the basis of objective reasons.

Against prevailing interpretations, I show that this line of thought plays an indispensable role in Kant's defense of free will against attacks from the theoretical standpoint. Kant's point is that someone occupying this standpoint must rely on her cognitive freedom from natural necessity: that is, she must take her beliefs about the natural causes of actions to be responsive to objective theoretical reasons, and here she cannot concede that the representation of causal necessity is merely the product of the workings of non-rational psychological processes. So the naturalistic perspective cannot expose freedom as an illusion. This provides the idea of free will with a robust invulnerability against an exclusively naturalistic worldview.

Table of Contents

Acknowledgements	ii
Introduction	iii-xii
Chapter I: The Metaphysical Commitments of Transcendental Idealism	2-28
Chapter II: Transcendental Idealism and Transcendental Freedom	29-50
Chapter III: Transcendental Freedom and Practical Freedom	51-66
Chapter IV: The Meaning of Practical 'Ought'	68-90
Chapter V: Naturalism and Normative Practical Authority	91-119
Chapter VI: Determinism and Objective Practical 'Ought'	120-139
Chapter VII: Responsibility and Possibility	140-162
Chapter VIII: The Normative and the Empirical Standpoint	164-202
Chapter IX: Freedom of Empirical Thought	203-228
Chapter X: Synthesis, Spontaneity, and Normativity	229-270
Chapter XI: Two Types of Epistemic Norms	271-295
Chapter XII: Empirical Belief and the Causality of Nature	296-323
Chapter XIII: Morality as the Ratio Cognoscendi of Practical Freedom	324-350
Bibliography	351-368

Acknowledgements

I gratefully acknowledge the generous financial support of the following institutions and agencies: the Department of Philosophy at UC Berkeley; the Graduate Division at UC Berkeley; and the Mabelle MacLeod Lewis Memorial Foundation. This dissertation project could not have been completed without their assistance.

The Department of Philosophy at UC Berkeley on the whole proved to be a good place to study philosophy and to think about Kant, freedom and normativity. My philosophical thinking and writing has benefited from countless discussions with faculty, graduate students, and undergraduate students. For stimulating exchange of ideas, I am grateful to (among others): Randall Amano, Joseph Barnes, Stephanie Basakis, Justin Bledin, Zack Bruce, Lara Buchak, Stanley Chen, Eugene Chislenko, Klaus Corcilius, Andy Engen, Peter Epstein, Dorothea Frede, Kinch Hoekstra, Niko Kolodny, Mike Martin, Erich Matthes, Julia Nefsky, Matthew Parrott, Kirsten Pickering, Mike Rieppel, John Schwenkler, Janum Sethi, George Tsai, and audiences at the Richard Wollheim Society and the Undergraduate Philosophy Forum. Special thanks are due, for very helpful feedback on various drafts and chapters, to Stephanie Basakis, Zack Bruce, Mike Rieppel, and Janum Sethi.

Outside of UC Berkeley, I am grateful to the members of the philosophy departments at Birkbeck College, London and the University of Tennessee, as well as the participants in Marcus Willaschek's Doktorandenkolloquium at the University of Frankfurt, for giving me the opportunity to present my ideas and for stimulating discussions. I am especially grateful to E.J. Coffman, Adam Cureton, and Marcus Willaschek for very helpful feedback on several dissertation chapters.

I was very fortunate to have Hannah Ginsborg and Jay Wallace as co-chairs of my dissertation committee; I could not have asked for a better pair of advisors. Without their constant feedback (which always struck a fine balance between encouragement and a healthy dose of skepticism), I would not have been able to develop and complete this project as I did. I am very much indebted for their reliability, patience, and above all for their numerous philosophical insights that helped improve my thinking and writing.

While not officially on the committee, Daniel Warren also was an unending stream of helpful feedback and encouragement; our many conversations through the years were a highlight of my philosophical education.

David Lynaugh deserves special mention for his invaluable practical advice and humor.

Most of all, I am grateful for their emotional support and love to Stephanie Basakis and to my parents, Gertrud and Walter Kohl.

Introduction

1: Aims, methodology and terminology of the dissertation

This dissertation is about Kant's theory of freedom or spontaneity. It is an attempt to work out: (1) what his doctrine of freedom asserts or entails; (2) his reasons for thinking that only a specific (incompatibilist) kind of freedom is one worth having; and (3) his reasons for thinking that we *are* free in his favored sense. The primary point of the dissertation is an interpretive one: its goal is to accurately portray Kant's views. But I do not believe that there is a *categorical* difference between doing interpretive, historical work and doing philosophy. After all, in doing history of philosophy one presupposes that one is dealing with a *great* past thinker, and this implies that that thinker had original, fruitful ideas. So in trying to accurately represent what these ideas are, one is obligated to search for *plausible and interesting* strands of argument. And in deeming a certain line of thought plausible and interesting, one is committed to thinking that it withstands at least certain kinds of critical scrutiny. So my dissertation is, at least with regard to (2) and (3), also committed to *assessing* Kant's conception of freedom in the light of relevant contemporary ideas and objections.

I want to make explicit two methodological presumptions that are central to my engagement with Kant's texts in this dissertation. First, I rarely use material from Kant's unpublished writings (in particular from the *Vorarbeiten*, *Reflexionen*, and lectures notes) to argue for an interpretive view; I do so only if some statement in these notes illustrates a point to which Kant is rather uncontroversially committed in view of his published writings. This is because I think that attempts to argue a controversial point by relying heavily on the unpublished writings face major interpretive difficulties.¹

A second central aspect of my approach to Kant's theory of freedom is that I do not ignore or bypass his epistemological and metaphysical views, in particular his transcendental idealism. The widespread strategy of bracketing the tenets of Kant's theoretical philosophy when discussing his practical philosophy is certainly understandable in many respects; but it has led to (what I consider) interpretive distortions, some of which I shall try to uncover in the dissertation. Kant himself undoubtedly considers his idealism a necessary presupposition of the cogency of his views of freedom, morality, and responsibility, and this is a point that someone trying to get clear about *what these views are* is obligated to take at face value. Now, since the doctrine of transcendental idealism plays a central role in my dissertation, it is worth clarifying right from the start my general attitude toward this doctrine. I am not convinced that transcendental idealism is correct², but neither do I consider it an incoherent or absurd doctrine. Moreover, I believe that Kant is right to think that the *realist* view he opposes (which holds that we know

¹ To mention only two problems here: a great many of the *Reflexionen* and *Vorarbeiten* show Kant laboring toward a view that he *may eventually* hold, but this does not justify identifying a particular position that he sketches in this process as his considered view. Likewise, with regard to the lecture notes, one would have to know the context of the relevant remark in order to confidently assert to what extent it reflects Kant's own position (he might be describing a view, such as Baumgarten's, that is not his own). For a good discussion of these and related methodological difficulties that afflict uncritical reliance on the unpublished writings, see Bojanowski 2006, pp. 186-90, and Hinske 1995.

² This is in part because it is not always crystal clear what transcendental idealism entails. But more importantly, I (pace, say, Allison) consider transcendental idealism to be a doctrine that has very substantive metaphysical commitments (cf. chapter I). I find it daunting to imagine what it would take to be consistently convinced that a grand metaphysical doctrine of this kind is correct.

objects as they are 'in themselves', in complete independence of their relation to our cognitive capacities) faces very significant difficulties. Therefore, the fact that Kant's theory of freedom presupposes his idealism does not (in my view) by itself suffice to disqualify this theory.

I now want to clarify an important terminological point. My dissertation frequently contrasts Kant's account with the views of *a naturalist opponent*. Since 'naturalism' is one of the most ambiguous terms in philosophy, I want to clarify what, exactly, I take this term to denote. The kind of naturalism I am concerned with affirms the following theses. (The characterization I offer is somewhat redundant, but since it has mainly illustrative purposes this does not matter.)

- (a) The only capacities that human beings possess are empirical powers: i.e., powers whose existence can be posited on the basis of empirical, scientific evidence.
- (b) Everything that happens takes place in accordance with laws of nature, and there are no causal powers that do not operate in accordance with laws of nature.
- (c) The only real, existent, efficacious causes of actions are those that figure in the explanations of the natural sciences.
- (d) Causes necessitate their effects; causal laws are deterministic.

Claims (a)-(c) embody the metaphysical position that all reality is exhausted by nature³, and the epistemological ('empiricist') commitment that the only legitimate evidence for ascribing properties to objects is based on sense-perception. That Kant takes his naturalist opponent to be committed to (d) reflects his own view that natural causes must necessitate their effects. When Kant hypothetically considers what would follow "if all causality were mere nature", he has in mind a scenario which comprises (a)-(d).⁴

My definition of naturalism here is not arbitrary: many contemporary philosophers who self-identify as naturalists would accept at least (a) and (c). But it is certainly stipulative in that it is intended to answer to my purpose of tracking the position that Kant himself was concerned about. Here two points must be emphasized. First, Kant does *not* picture his naturalistic opponent as essentially *materialistic*.⁵ And this point has a very important consequence: Kant's worries about the implications of naturalism *never* relate to the anxiety that naturalism would imply that there are no mental phenomena or that our deliberation and choice lack causal efficacy. For Kant, the naturalist is fully entitled to the view that mental representations are real natural states that have causal efficacy; and he accepts that among the natural sciences mentioned by (c), empirical psychology has an irreducible role and place (even though he doubts its scientific potential).

Second, the version of naturalism that Kant is concerned with is somewhat anachronistic, since some contemporary naturalists would reject (d) on the basis of the implications of quantum

³ See Kim 2003.

⁴ See CPR, A 534/B 562. Kant's naturalist or empiricist opponent adopts the antithesis of all the 'Antinomies of Pure Reason'. The view that comprises (a)-(d), appears in Kant as *one* version of *transcendental realism*. Note that Kant himself uses the term 'naturalism' in a very different and idiosyncratic sense (CPR, A 855-6/B 883-4). Moreover, it must be admitted that the term 'empiricism' carries lots of different connotations in Kant; see footnote 85 in chapter II below for some discussion.

⁵ This is because Kant himself does not identify natural causes with physical causes. I argue for this point in chapter II, section 2. Hatfield 1990, pp. 16-7 distinguishes between a metaphysical and a methodological version of naturalism; he stipulates that a metaphysical version of naturalism must be materialistic, and that a methodological version of naturalism is committed to explaining the mind in terms of non-physical natural laws (Hume is a methodological naturalist in this sense). This is a bit puzzling because a 'methodological' naturalist typically shares the 'metaphysical' naturalist's *ontological* commitment to the non-existence of non-natural causes, objects and properties.

mechanics.⁶ Nonetheless, in my dissertation I mostly operate under the assumption that natural causes are deterministic, because Kant himself considered the thesis that natural causes necessitate their effects an a priori truth. It is worth noticing that there is much controversy about the extent to which quantum mechanics really 'disproves' determinism⁷; likewise, some hold that the claim that *macroscopic* processes are deterministic is not endangered by the implications of quantum mechanics.⁸ Apart from these issues, I share the generally accepted view that for someone with incompatibilist intuitions, the recourse to indeterminism is fruitless⁹, and I shall on occasion suggest how Kant's incompatibilist arguments might be adjusted to the assumption that natural causes operate in an indeterministic manner.

2: Overall claims and structure of the dissertation

The most important claim that I shall try to establish in the course of the dissertation is, in very programmatic terms, that for Kant absolute metaphysical freedom from 'the causality of nature' is a condition of the possibility of both practical and epistemic normativity. Kant holds that our ability to put forward ('legislate') normative principles, and to act under the governance of such principles, requires that we are not exclusively determined by natural laws and forces. This point applies to principles of both practical and epistemic rationality, and to both practical and cognitive acts. Most strikingly, Kant holds that the principle that natural events are governed by nomological, empirical necessity itself is fundamentally a cognitive norm for object-directed thought (even if it is also – via Kant's idealism – a condition on the possibility of *objects* of experience).¹⁰ He claims that the objective validity of this principle presupposes absolute metaphysical freedom of thought – to put the point (again) in very programmatic terms, we can only impose causality on nature if we are free from natural causality. The 'master argument' that I shall attribute to Kant is that our commitment to the idea that we have practical freedom (of will) from nature enjoys a certain robust kind of metaphysical and epistemological invulnerability against standard naturalist/empiricist attacks, because the naturalist herself must presuppose freedom of (empirical) thought from nature in advancing her causal, naturalistic explanations.

I now want to describe the overall structure of my dissertation and to sum up the central points and arguments of each individual chapter. The dissertation divides into three parts.

In part 1, I discuss the basic conceptual, metaphysical and epistemological foundations of Kant's theory. As I pointed out, I take seriously Kant's insistence that accepting his idealism is a

⁶ Some might also reject (b) because they believe in 'singular' causation (which does not presuppose natural laws).

⁷ For the view that determinism is still very much a live issue, see Earman 2004.

⁸ See, for instance, Baumgartner 2008.

⁹ See Loewer 1996.

¹⁰ In tracing Kant's conception of epistemic normativity to the role of the pure concepts of the understanding, such as causality (as well as those of theoretical reason, such as unity), my approach to the contrast between the 'natural' and the 'normative' in Kant differs from Hatfield's 1990. Hatfield discusses the extent to which a 'naturalization' of processes of *spatial representation* and of judgments concerning the spatial properties of objects (pp. 6-7; p. 16) is possible without appeal to irreducibly normative elements, and thereby places his main emphasis on Kant's doctrine of *sensibility* (p. 7). More generally, Hatfield discusses whether mental states can be reduced naturalistically, without invoking normative notions such as 'true', 'justified' or 'valid' (p. 15). I am less concerned with an analysis of mental states or content; my main focus is on whether thinking, reasoning and judging about the natural world stands in need of epistemic *oughts* and, if so, what this implies about our self-conception as free thinkers. Of course, this *has* implications for the question of whether we can understand judgment or belief, including spatial judgment and belief, without appeal to normative considerations.

precondition for accepting his views about freedom. In **Chapter I**, I sketch my conception of the general tenets of Kant's idealism. I argue for three claims. First, the relation between *appearances* and *things in themselves* is intended by Kant to be one of numerical identity, at least with respect to beings who can be considered free and rational. Secondly, the distinction cannot be ontologically neutralized: Kant is committed to the view that objects of our knowledge *have* a character (Beschaffenheit) that *differs* from the (empirical, phenomenal) character they (*those same objects*) possess by virtue of being related to our cognitive faculties. Third, while we have no theoretical insight into the character that things have in themselves, we do have conceptual resources for ascribing an indeterminate noumenal constitution to objects and even for characterizing this constitution positively (albeit not for the purposes of cognition). My interpretation shows that neither the view that Kant's idealism implies a kind of substance dualism (the 'two-world view') nor the view that Kant's idealism is devoid of all metaphysical implications (the 'merely epistemological' 'two-aspect view') are correct – we should adopt a kind of middle ground between the two extremes.

In **Chapter II**, I explain Kant's conception of *transcendental freedom*. This conception has seemed mysterious to many philosophers and has provoked irreconcilably different (compatibilist, incompatibilist, hybrid) interpretations. Against the background of the results established in chapter I, I argue for two points. First, I show that attempts to provide a metaphysical deflation or a compatibilist interpretation of Kant's conception of freedom cannot succeed. We must read Kant as holding that we are, *as a matter of ontological fact*, free from causal necessity. This point is to be understood in the context of Kant's idealism: we can consider ourselves *both* as free from causal necessity *and* as subject to such necessity insofar as we can consider ourselves both as noumena and as phenomena. Secondly, I show that Allen Wood's famous idea that Kant tries to show "the compatibility of compatibilism and incompatibilism" is misleading; I suggest that we should rather understand Kant as trying to show the compatibility of hard determinism and libertarianism. If we consider ourselves and our actions as natural phenomena, from the empirical standpoint, we *must not* consider ourselves as free. We can legitimately consider ourselves and our actions as free insofar as we can consider ourselves, from the noumenal standpoint, as standing outside of the deterministic order.

Kant's theory of freedom, as I interpret it, implies that we can attribute our actions *both* to the deterministic, temporal causality of nature *and* to the non-deterministic, atemporal, spontaneous causality of freedom. Kant is happy to concede that this picture is theoretically elusive (albeit coherent). The belief in our freedom from causal necessity is theoretically fruitless *because* the idea of freedom cannot be instantiated in our experience of nature. This does not bother Kant because his idealism makes systematic ontological space for capacities of objects that differ from the capacities objects possess *qua* natural, spatiotemporal objects of human cognition, and that are *therefore* beyond the limits of human cognition. And his standpoint distinction suggests that we employ the concept of freedom not for the purposes of theorizing (observing, explaining, predicting) but for the purposes of answering questions about what *ought* to happen. In this manner, Kant's theory of freedom accommodates the naturalistic complaint about standard libertarian accounts of freedom: Kant differs from standard libertarianism in that he refrains from providing a conception of freedom that intends to rival the naturalistic, deterministic model of event-event causation *on explanatory grounds*.

This provides my answer to the first of the aforementioned questions: (1) What does Kant's doctrine of freedom assert or entail? No attempt to render the concept of free human

agency theoretically determinate or tractable, and no metaphysical theorizing about 'how the causality of freedom works', is to be expected from Kant or from me.

However, I go on to ask whether we can be *satisfied* with this conception of freedom. I suggest that in order to be convinced that Kant's conception of freedom is worth paying the price of theoretical elusiveness, we would have to be convinced, first, that his conception of freedom is one, or indeed the *only* one, worth having; and second, that we have reason to believe that we *are* free in his ambitious sense. This raises the the aforementioned questions (2) and (3): Why does Kant think that compatibilist views of freedom are unsatisfactory? What reason do we have to believe in a kind of freedom that we cannot know through observation or experience? These two questions set the agenda for the remainder of the dissertation.

Before addressing these issues, I consider, in **Chapter III**, the relation between transcendental and practical freedom. I argue that practical freedom is a specific kind of transcendental freedom, namely, transcendental freedom as instantiated by the faculty of *will*. This means that Kant's conception of practical freedom is as metaphysically ambitious, incompatibilist, and theoretically elusive as his conception of transcendental freedom. The major obstacle for this interpretation is Kant's discussion of freedom in the Canon of Pure Reason, which seems to suggest a metaphysically deflationary, compatibilist treatment of practical freedom. I argue that we need not and should not understand the Canon passage in this manner. However, although I claim that Kant's treatment of practical freedom in the Canon is compatible with his incompatibilist treatment of practical freedom in the Dialectic, I also suggest that the discussion in the Canon is unsatisfying and falls short of Kant's considered critical views.

In the appendix to chapter III, I offer a systematic taxonomy of the various concepts of freedom (transcendental, practical, positive, negative,...) that Kant employs throughout his 'critical' writings. This concludes the first part of my dissertation.

The second part of my dissertation addresses the aforementioned question (2): why does Kant think that compatibilist views of practical freedom are unsatisfactory? What are the grounds of his stark and uncompromising incompatibilism about practical freedom?

Chapter IV begins my discussion of Kant's incompatibilism by focusing on his claim that if naturalism (as defined above) were true, practical ought judgments would be meaningless. To understand why he thinks this, we must first understand how he conceives of the meaning of practical oughts; this is the topic of chapter IV. I distinguish between the meaning of *categorical* and *hypothetical* practical judgments (CPJ, HPJ). I show that while understanding the meaning of CPJ is relatively straightforward, Kant's conception of the point and meaning of HPJ is very hard to interpret. This is because his texts suggest both a *Humean* and an *anti-Humean* understanding of HPJ. The Humean view is that practical reason is uninvolved in establishing HPJ and in non-moral motivation. The anti-Humean view is that valid HPJ and non-moral motivation require a contribution of practical reason that is irreducible to the contribution of desire. I propose an interpretation that combines Humean and anti-Humean elements: what an agent has non-moral reason to do is not independent of her contingent desires as such but from her *transient, fleeting* desires.

On the basis of my interpretation of Kant's conception of the meaning of practical ought judgments, I suggest that his claim that such judgments would be meaningless 'if all causality were mere nature' divides into two conceptually separable worries. First, there is a worry about the *content* of practical ought judgments: such judgments, in deeming courses of action *required*,

permitted, forbidden, advisable or foolish, purport to have a special authority that they would lack if they had to be attributed to the influence of contingent natural causes. Second, there is a worry about the *formula* of practical ought judgments: the formula 'ought' signals that the *addressee* of the judgment has the freedom to comply with or to violate the norm that the judgment represents; and if all causality were mere nature, no agent would ever have the freedom to go 'either way' with regard to a practical norm.

Chapter V provides a detailed discussion of the first of these worries. I begin by arguing that there is strong intuitive support for Kant's point that if all our practical reasoning and judging were the causal upshot of contingent natural conditions, our judgments would fail to have *objective, necessary* validity and authority. There could not be *categorical* imperatives. But this point does not support a worry about practical normativity *as such*. Naturalists about normativity do not believe in categorical imperatives, but they (or at least many of them) hold that there *are* truths about what agents have reason to do which depend on facts about the contingent desires of agents. Given this naturalistic, 'desire-based' conception of practical normativity, it cannot count against practical judgments which represent desire-based reasons that they causally depend on contingent desires. However, I argue that desire-based theories fail as an analysis of practical normativity because they lack resources for distinguishing between desires that have and desires that lack normative authority; and thus they unwittingly imply that every agent *always* does what she has reason to do. From the rejection of desire-based theories, it follows that even the non-moral practical reasons represented by non-categorical practical judgments are irreducible to contingent desires. Hence, if such judgments were caused by contingent desires, they would lack normative authority over the desires that happen to affect an agent at a given point in time (i.e., they could not deem acting on any given contingent desire *foolish* or *advisable*). So the content of both CPJ and HPJ is incompatible with the assumption that our practical reasoning and judging is the causal upshot of contingent natural states.

Chapter VI provides a detailed discussion of the second worry that I sketched in chapter IV. I begin by discussing the prevalent, compatibilist view of how Kant solves the problem determinism poses for human agency: on this view, all the standpoint of agency requires is that the agent has *epistemic freedom of choice* (i.e., she *can* choose various things *for all she can know*), and Kant shows that determinism does not threaten this kind of freedom. I argue that this interpretation is textually indefensible. I then expound Kant's real worry: namely, the worry that if our choices were exclusively subject to causal necessity, we would never have the freedom to go either way with regard to a practical norm. This would mean that practical *ought* judgments are deprived of their point and meaning. One controversial aspect of this argument is the idea that a human agent can never be *rationally necessitated* to choose the right thing: if she is necessitated to choose the right thing, she is subject to the non-rational influence of material, natural causality. I show that this radical conception of the imperfection in our practical rationality is plausible: it rests on the view that every human agent is always subject to desires that impel her to choose in ways that resist governance by desire-independent moral reasons. This suggests an interesting conception of moral psychology whose plausibility is independent of Kant's commitment to hedonism about non-moral motivation.

In **chapter VII**, I discuss Kant's incompatibilism about *moral responsibility*. I attribute a straightforward incompatibilist argument to Kant: holding an agent responsible requires that the agent could have acted otherwise, and 'if all causality were mere nature' no agent could ever act differently than she in fact does. However, Kant affirms this 'principle of alternative possibilities'

only with regard to *inner* action (choice): judgments of responsibility concern the quality of the agent's will which depends on the agent's acceptance or rejection of the reasons represented by moral oughts; but whether or not an agent acknowledges the weight of moral reasons is only imperfectly correlated with her observable ('outer') actions. (Our outer actions are 'morally contingent'.) I show that Kant's argument can be defended against contemporary compatibilist objections. I then go on to raise a major interpretive difficulty that has so far gone unnoticed. The difficulty is this: Kant's incompatibilist argument trades on the idea that in a deterministic world, alternative actions (choices) would not be (in his technical sense) *really* possible ('real possibility' being defined by reference to what the laws of nature allow). But his own libertarian position is *not* entitled to the view that alternative actions are 'really possible' *either*: Kant himself can only say that alternative actions are *logically possible* (conceivable) – and surely his compatibilist opponent can say exactly the same. I suggest that this difficulty can be resolved by noting that Kant and his naturalist opponent interpret the concept of 'mere logical possibility' in different ways. For Kant, the claim that something is contrary to the laws of nature, and hence *merely logically possible*, does not imply that it is impossible in an absolute ontological sense, for on his view the laws of nature constrain only our *empirical* capacities. But Kant's naturalist opponent thinks that our empirical capacities exhaust our powers; hence she must accept that whatever is ruled out by the laws of nature (and merely conceivable) is beyond our power.

This concludes the second part of my dissertation. The result of my discussion here is that Kant has a range of interesting, original, and defensible arguments against a compatibilist conception of freedom of will which jointly suggest that his starkly incompatibilist conception of freedom might be worth paying the price of theoretical fruitlessness.

Part 3 of my dissertation addresses the aforementioned question (3): what reason do we have to believe that we are practically free? Since I take Kant's answer to this question to be fairly complicated, my attempt to represent his answer involves a rather long but illuminating detour.

In **chapter VIII**, I discuss in detail Kant's *standpoint distinction*. This distinction is often taken to contribute to Kant's defense of our practical freedom, but it is mostly those who seek to *metaphysically deflate* Kant's theory of freedom who avail themselves of Kant's appeal to standpoints. As I argue in chapter II, such deflationary interpretations are untenable. Yet Kant often does seem to defend his conception of freedom by invoking a distinction between standpoints. So what is the point of this distinction?

I show that the distinction can be drawn in multiple ways. First, it can be drawn by appeal to the conceptual resources employed from the two standpoints: *merely human* concepts (that involve a reference to merely human forms of sensibility) versus *more than human* concepts (that would be shared by all thinking or rational beings). Second, it can be drawn by appeal to the point and purpose associated with the two standpoints. The merely human standpoint is the standpoint of empirical cognition that attempts to explain, predict, and observe events and objects in the spatiotemporal, natural world. The common view is that this standpoint is to be contrasted with a *practical* standpoint from which one establishes what one *should choose*. However, I show that Kant's conception of the non-empirical standpoint is broader: it is a normative perspective that considers people *not merely* as free subjects of volition (asking how they ought to choose) *but also* as free subjects of cognition (asking how they ought to think). This standpoint qualifies as a 'more than human' perspective on things because it represents human beings in terms of the 'more than human' concept of freedom (or spontaneity).

Understanding Kant's standpoint distinction in this unorthodox manner raises two questions that turn out to be crucial for the further course of my enquiry. First, *what kind of freedom* is exhibited by the subject of empirical *belief*, and how does it compare to the freedom exhibited by the subject of *choice*? Second, what are the *epistemic norms (oughts)* that are addressed, from the normative standpoint, to the subject of empirical thought? I take up these questions in subsequent chapters.

I continue my discussion of Kant's standpoint distinction by arguing for a number of claims: first, that the two perspectives are necessarily distinct from one another; secondly, that they are interdependent (i.e., each perspective depends on the claims established from the other perspective); and third, that they are in harmony with each other (i.e., their respective conceptions of human beings are consistent). With respect to this last issue, I show that Kant's attempt to achieve a harmony between the two standpoints essentially presupposes his transcendental idealism: it is only through accepting the phenomena/noumena distinction, and through rejecting the idea that determinism has the status of a universal metaphysical truth, that Kant can render the two claims, 'we are free' and 'we are causally determined' consistent. But although the appeal to standpoints cannot by itself (i.e., without the help of Kant's idealism) defend our self-conception as free agents, it does have *some* legitimizing force: since the concept of freedom is, for Kant, theoretically fruitless, its legitimacy essentially depends on the idea that there is a legitimate perspective on things that is not concerned with theoretical (e.g., explanatory) questions. This is the *normative* standpoint, *from which alone* the theoretically fruitless concept of freedom may be fruitfully and legitimately used.

I conclude my discussion of Kant's standpoint distinction by comparing it to, and contrasting it with, 'Kantian' standpoint distinctions in contemporary philosophy. I show that insisting on the differences between Kant's own account and contemporary Kantian accounts is motivated not just by philological pedantry, but by deep and interesting philosophical differences between Kant and (e.g.) Christine Korsgaard or Richard Moran.

In chapter IX, I take up one of the questions raised by my interpretation of Kant's standpoint distinction: *what is freedom of empirical thought/belief*? I argue that for Kant, our theoretical freedom consists in the fact that our cognitive responses to given perceptions are neither caused by external objects nor by contingent psychological processes. Rather, our beliefs result from our reflective conception of what we ought to believe. While this suggests a kind of 'freedom as orthonomy' (a freedom to believe the *right* thing), I show that Kant considers the *imperfection* inherent in our rationality essential to our self-conception as free thinkers: our freedom of thought involves a susceptibility to come under the influence of *optical and transcendental illusions*. The structural disanalogy between *practical* and *theoretical* irrationality is this: in the practical case, our understanding of how we ought to choose is – at a basic level – relatively immune against corruption, but our choices tend to fail to reflect this understanding; by contrast, while our beliefs typically *do* reflect our conception of what we ought to believe (we cannot, typically, believe *at will*), this conception is very amenable to being corrupted.

In chapter X, I discuss a different aspect of Kant's conception of epistemic agency: the role of spontaneity and normativity in the 'original synthesis of intuitions'. This chapter is bound to be among the most controversial pieces in my dissertation, for it contradicts various established ways of thinking about the role of theoretical spontaneity and normativity in Kant's theory of synthesis. I argue, on the basis of close textual analysis, for three central and related claims. First, I argue that the original synthesis of intuitions, at least the one that brings given

intuitions under the *dynamical* categories (such as 'substance' and 'causality') is not a *blind* but a *fully conscious* process; correspondingly, the rules that govern synthesis are reflected in the conscious awareness of the *subject who performs* synthesis rather than (only) in third-personal theorizing about this subject. Second, I argue that the rules governing synthesis have the status of *norms* that prescribe how we ought to think about nature (at a very basic, conceptually indeterminate level), and that these normative rules are the principles of the understanding. This has the upshot that the principle, 'Every event has a cause', which appears to be about *objects* in the world, is *also* a rule for *subjects* who have experience of the objects that the principle refers to. While this may seem like an odd idea, I suggest that it is strictly required by Kant's 'Copernican Revolution' and his claim that "the conditions of the possibility of experience as such are *at the same time* conditions of the possibility of the objects of experience in general (*überhaupt*)". The causal principle is a condition on objective events *and at the same time* on objective experience of events; and it owes its status as an ontological truth entirely to its status as a rule that governs objective experience. Third, I argue that the subject of synthesis has a genuine freedom to violate the rules of synthesis: this freedom, I suggest, derives from our susceptibility to come under the influence of transcendental illusion.

An important upshot of my argument here is that it gives a *partial* answer to a question that came up in my interpretation of Kant's standpoint distinction: *what are the epistemic norms* that the normative standpoint addresses to the subject of empirical thought? My interpretation of the synthesis of intuitions suggests that these norms include the principles of the understanding as norms that govern our thinking about nature at a basic level; a thinking that Kant calls experience in general (*überhaupt*).

In **chapter XI**, I show why this answer is only partial. The norms of empirical cognition include not only the *constitutive principles of the understanding* (such as the principle that every event has a cause) but also the *regulative principles of theoretical reason* (such as the principle that we ought to strive for systematic unity in our thinking about nature). My understanding of the contrast between constitutive and regulative principles as one between *two kinds of epistemic norms* is motivated by the fact that Kant thinks that these two principles jointly yield necessary and sufficient conditions (or criteria) of empirical truth. The constitutive principles are *necessary but insufficient* conditions for empirical truth insofar as compliance with these principles is required for experience *in general*, i.e., for object-directed thought at a basic level that falls short of conceptually determinate cognition. Compliance with the regulative principles allows for transgression from experience *in general* to experience qua conceptually determinate cognition of objects (through specific empirical concepts). I show that Kant's appeal to the regulative ideal of systematic unity (of empirical laws and concepts) as a criterion for empirical truth does not commit him to a questionable form of coherentism about empirical knowledge: the claim that our pursuit of empirical truth requires the normative guidance of the ideal of systematic unity does not preempt the need for a receptive, perceptual constraint on empirical thought.

In **chapter XII**, I examine the grounds of Kant's incompatibilism about freedom of empirical thought. I start with an idea advanced by Thomas Nagel and David Wiggins, who suggest that incompatibilist worries in the theoretical domain can be defused by the idea that our beliefs are ideally 'determined by the truth'. I argue that this idea reflects a naïve empiricism that cannot be sustained. Our beliefs about the world are not simply dictated to us by the impact that external objects have on our sense organs. I discuss various ways in which empirical belief formation calls for an active contribution of the subject of cognition, and I argue that this

contribution cannot happily be understood as being controlled by the contingent mechanisms of the empirical mind. The centerpiece of my discussion here concerns the status of the concept of *causation*. I argue that Kant's conception of the spontaneity of mind as requiring independence from natural causes is centrally provoked by Hume's original skepticism: our conception of causation is underdetermined by the content of our sense perceptions, so *if nature forces us* to think in causal terms, the 'force' comes from within – from the contingent natural 'habits' of the mind, in which case the concept of causality loses its claim to objective validity. I show that this worry about the origin and status of the concept of causality is not preempted by contemporary discussions of causation, and that Kant's conception of absolute theoretical spontaneity as (part of) his response to Hume is of continuing interest and relevance.

These considerations put me, finally, in a position to answer the aforementioned question (3): what, on Kant's account, justifies the belief in freedom of will? In **chapter XIII**, I argue that Kant's defense of free will crucially depends on an appeal to freedom of thought. In the Critique of Practical Reason, Kant seeks to infer our freedom of will from our awareness of the moral law (as a 'fact of reason'). I argue that given the reciprocal relation between freedom and morality, this inference would be unacceptably dogmatic *unless* Kant could somehow back up the idea that we can take our awareness of practical necessity at face value (i.e., that we can consider it as genuine rather than a mere 'figment of the imagination'). On my interpretation, the appeal to freedom of empirical thought backs up the appeal to the moral law as a fact of (practical) reason. To correctly understand the argumentative role of the appeal to theoretical freedom we must, I argue, make explicit the dialectical context in which Kant brings up the idea of freedom of thought. This dialectical context is generated by the following question: why would we have reason to doubt that we can take our awareness of practical necessity – in the form of the moral law – at face value as a *ratio cognoscendi* of the freedom of will (the *ratio essendi* of the moral law)? The answer seems to be that such doubts derive from the naturalistic view that there are no necessary practical rules, that all concepts are contingent, and that all thinking and judging is conditioned by contingent natural causes. The point of Kant's appeal to freedom of empirical thought is that this naturalistic picture is incoherent, because it presupposes the legitimacy and objective validity of the concept of causality, which – according to Kant – itself requires the absolute spontaneity of the understanding.

I show that the appeal to freedom of thought plays a double role in the defense of freedom of will. First, it preempts the naturalistic or deterministic attack on the possibility of freedom of will. Second, it justifies the inference from our awareness of practical necessity to our freedom of practical reason by providing us with a *paradigm case* of the reliability of this *type* of inference. This paradigm case is our inference from our awareness of the epistemic necessity of the causal principle to the spontaneity of the faculty that must be considered the source of this principle (i.e., the understanding). This appeal to freedom of thought does not strictly prove that we have freedom of will, but it does put the inference from our awareness of practical necessity to our freedom of will (or pure practical reason) beyond reasonable doubt.

Hence, the autonomy of the understanding can be invoked in defense of the autonomy of the will. The result is a unified view of ourselves as free, autonomous thinkers and choosers.

Part 1
**The conceptual, epistemological, and metaphysical
framework of Kant's theory of freedom**

Chapter I: The Metaphysical Commitments of Transcendental Idealism

Kant repeatedly insists that endorsing his doctrine of transcendental idealism is necessary for "saving" the idea of freedom (CPR, B XXVI; A 536-7/B 564-5; GMS: 4, 456-7; KpV: 5, 94-5). Hence, an interpretation of Kant's theory of freedom must be based on an interpretation of Kant's idealism. Somewhat more specifically, one must at least provide an understanding of Kant's distinction between appearances (phenomena), the familiar spatiotemporal objects of our ordinary experience, and things in themselves (noumena)¹¹, the mysterious things that stand 'behind' appearances and that are not in space or in time: for Kant clearly thinks that this distinction is the key toward legitimizing our self-conception as free agents.

We can distinguish between two main, contrasting types of interpretations: *metaphysical* and *merely epistemological* ones. The former view affirms, whereas the latter view denies, that Kant's talk about things in themselves has ontological import. In this chapter, I shall develop a *moderately metaphysical* interpretation that occupies a kind of middle ground between standard metaphysical and merely epistemological readings. This will provide the general framework for the interpretation of Kant's theory of freedom that I will develop in subsequent chapters.

1: Allison versus Langton: making room for a 'third way'

I want to begin by looking at the dispute between two major proponents of the competing lines of interpretation. Rae Langton's metaphysical account runs as follows: when Kant talks about things in themselves, he refers to substances that are the bearers of 'intrinsic', non-relational properties; by contrast, when Kant talks about appearances, he refers to the relational properties of substances.¹² Henry Allison, the main proponent of a 'non-metaphysical' interpretation, concedes that Kant's talk about things in themselves involves an abstraction from *one particular* relation, namely, that of an object's relation to our sensibility; but he criticizes Langton for slipping from the abstraction from this particular relation to the abstraction from *all* the relational properties that objects might have.¹³ I agree with Allison here: even if it turned out that we must understand Kant's conception of things in themselves in terms of substances bearing inner properties (I shall address this issue in section 5 below), it seems unwarranted to dogmatically deny that things in themselves might also stand in certain relations to one another, some of which may ground the phenomenal relational features that constitute our knowledge of objects.¹⁴

¹¹ I shall treat 'noumena' and 'things in themselves', as well as 'phenomena' and 'appearances', as synonymous terms. This is not to deny that there may be interesting distinctions to be drawn here, but (as even Allison 2004, who firmly promotes the significance of these distinctions, admits; see p. 58) Kant tends to run these pairs of concepts together. Nothing of substance will hinge on this point for the purposes of this paper. For a systematic treatment of difference between the appearance qua 'undetermined object of empirical intuition' and qua phenomenon ('determined under empirical concepts'), see Longuenesse 2001, especially chapter 1.

¹² See Langton 1998, p. 50. Langton can be read as proposing that the contrast between phenomena and noumena concerns two ways of characterizing *one* substance (through relational or intrinsic features). This would be a metaphysical 'two-aspect view' (cf. Watkins 2004, p. 318). Allison seems to read her as endorsing a 'two-world view', i.e., a kind of dualism between *two separate substances*; he claims that his epistemological account refrains from Langton's idea to think of a thing in itself "as a separate *substantia noumenon*" (2004, p. 52).

¹³ See Allison 2004, p. 52 ff.

¹⁴ In particular, the relation of *transcendental affection* seems to be a relation among things in themselves. When Kant says that that the representation of outer objects requires that we are affected with sensible data that must pass through our sensible or receptive faculties (CPR, A26-7/B 42-3), he is, arguably, not referring to an empirical, observable relation between phenomena. Apart from the issue of transcendental affection, the idea that

Now, Allison seems to assume that rejecting Langton's account rules out *any* interpretation that can be deemed 'ontological'.¹⁵ He may be assuming this because the relation objects have to human sensibility is an epistemic relation: to abstract from this relation is to abstract from the conditions under which we can know things. Allison might hold that emphasizing the epistemic nature of this relation is the only way to avoid that Kant's abstraction from this particular relation (in his conception of noumena) becomes ad hoc; an 'ad hocness' that would encourage a generalization to all relations and thus lead to Langton's idea that Kant's conception of a noumenon requires abstracting from relational features as such.

I concede to Allison that Kant's abstraction from the relation between objects and our cognitive faculties (an abstraction which is the core of Kant's conception of noumena) *is* motivated by epistemic considerations, i.e., by the idea that our cognitive faculties impose conditions on our knowledge of objects. But this will support an anti-metaphysical interpretation of Kant's idealism only under the assumption that the relation that objects bear to our sensibility is exclusively an epistemic relation.¹⁶ This assumption ignores the possibility that our cognitive faculties determine not only the character of *our knowledge of objects* but also the character of the *objects of our knowledge*. That the relation in which objects stand to our cognitive faculties has this double, epistemic-ontological nature, is strongly suggested by the following remarks:

If intuition must conform to the constitution of the objects, I do not see how we could know anything of the latter a priori; but if the object (as object of the senses) must conform to the constitution of our faculty of intuition, I have no difficulty in conceiving such a possibility. (CPR, B XVII)

There are only two possible ways in which synthetic representations and their objects can establish connection...and, as it were, meet one another. Either the object alone must make the representation possible, or the representation alone must make the object possible. In the former case, this representation is only empirical, and the representation is never possible a priori. In the latter case...the representation is a priori determinant of the object, if it be the case that only through the representation is it possible to know anything as an object. (CPR, A 92/B 124-5)

...the conditions of the possibility of experience...are at the same time conditions of the possibility of the objects of experience. (CPR, A 158/B 197)

These quotes suggest that our cognitive faculties determine the (phenomenal) character of the objects of our (human) knowledge, and that it is (only) by virtue of possessing this phenomenal character that objects can be known by us. This implies that objects as they are in themselves, i.e., insofar as they do *not* 'conform to' the character of our cognitive capacities, have a certain noumenal character that is (precisely because of this lack of conformity to our cognitive faculties) unknowable to us. Metaphysics and epistemology are inseparably intertwined here. Now, all the rejection of Langton's interpretation implies is that we have no right to insist that the noumenal constitution of objects could *only* include non-relational properties. This allows: first (1) the idea that objects as they are in themselves can be characterized both in terms of relational features (other than the relation between objects and human sensibility) and in terms of 'intrinsic' (non-relational) properties¹⁷; and second (2), the idea that objects as they are in themselves can

one must appeal to relations among things in themselves to *ground* the spatiotemporal properties of phenomenal objects is put forward, in different forms, by Broad 1978, pp. 23-7; Findlay 1981; and Paton 1970, p. 417 ff.

¹⁵ See Allison 2004, p. 52: "Let us call the former [Langton's] the ontological and the latter [his own] the epistemological sense of a consideration of something as it is in itself."

¹⁶ This assumption is in play not only in Allison but also in Hudson 1994, pp. 28-30.

¹⁷ This proposal is advanced by Warren 2001 and unpublished. Warren argues further that the inner determinations of things in themselves are explanatorily prior to their relational features.

be characterized as having *some* properties other than those that they possess qua objects of human, spatiotemporal intuition. (2) remains agnostic about what the metaphysical nature of those non-spatiotemporal properties might be (i.e., about whether they are intrinsic or relational).

It is clear from Allison's criticism of Langton that he would consider it illegitimate to claim that things in themselves are to be characterized in terms of (1) both relational and intrinsic properties. This claim, like Langton's interpretation, seems "minimally informative about the *nature*" of things in themselves¹⁸, and is hence, for Allison, to be rejected as an interpretation of Kant's epistemically humble notion of a thing in itself. But this criticism does not affect (2): the suggestion that objects in themselves have non-spatiotemporal properties is not informative at all, because it does not aspire to say anything about the metaphysical nature of those properties.

Allison might object that *even* this 'minimal' suggestion fails to do justice to Kant's epistemic humility. Kant famously affirms the unknowability of things in themselves: how, then, could he claim to know that things in themselves are *not* spatiotemporal? I think that Kant holds that saying what an object is *not* is compatible with the thesis of unknowability, because negative determinations do not yield 'proper knowledge' of what objects are like in themselves:

If we suppose an object of a non-sensible intuition to be given, we can indeed represent it through all the predicates which are implied in the presupposition that it has none of the characteristics proper to sensible intuition; that it is not extended or in space, that its duration is not a time, that no change...is to be met with in it, etc. But there is no proper knowledge if I thus merely indicate what the intuition of an object is not, without being able to say what it is that is contained in the intuition. (CPR, B 149; see also A 358-9)

The idea expressed in this passage is, I think, implicit in many contexts where Kant puts forward the unknowability thesis. He often moves (see, for instance, CPR, A 42/B 59) directly from the claim that objects as they are in themselves are not in space and time to the claim that what they are in themselves remains unknown to us. The only way to rationalize this move is by supplying the premise that saying what an object is not does not suffice for proper knowledge of the object. This epistemic premise is linked to a metaphysical tenet: Kant holds that "reality is something, negation is nothing", a "nihil privativum" (CPR, A 291/B 347). So in saying that objects as they are in themselves do not have spatiotemporal properties, we fall short of cognizing any ontologically real features and hence fall short of achieving 'proper knowledge'.¹⁹

¹⁸ Allison 2004, p. 18.

¹⁹ Of course, one may doubt whether this suffices to defuse Jacobi's venerable charge that Kant is inconsistent in proclaiming that unknowable things in themselves lack spatiotemporal properties. (Although I find Guyer's 1987, pp. 335-6 argument that this charge rests on a misunderstanding fairly persuasive.) One thing that I would here is that the unknowability thesis pertains to the impossibility of substantive, synthetic a priori knowledge of things in themselves: we cannot know, through mere concepts, non-trivial truths about what non-sensible objects must be like (CPR, B 73). Now, the claim that objects in themselves are not spatiotemporal is an analytic claim that merely makes explicit what is contained in the concept of an object of a non-sensible intuition (CPR, B 149). Hence, this claim can be considered as both trivial and (because of its negativity) as failing to determine ontologically real features. My interpretation here strongly conflicts with Hogan's 2009 argument that Kant's unknowability thesis itself rests on a *positive* claim about noumenal reality, namely, on the claim that noumenal reality contains features that lack a determining ground that would be required for (non-divine) knowledge of noumena. I agree with Hogan that Kant is committed to the claim that noumenal reality lacks the determining grounds required for our human theoretical cognition. But our only cognitively significant notion of a determining ground requires a schematized notion of causality, i.e., a notion of a ground-consequent relation that is applicable to objects in space and time. Things in themselves are not in space and time. Hence, the notion of a determining ground that beings like us require for cognition is in principle inapplicable to noumena. But this

What I have argued thus far provides the sketch of a modest metaphysical interpretation. It reads Kant as being committed to the idea that objects have a constitution in themselves that differs from the constitution they have as objects of human sensibility and that we cannot positively determine on theoretical grounds. This reading is ontological but eschews even minimally determinate theories about what kinds of properties things in themselves exhibit.

Now, some proponents of a metaphysical interpretation – those who favor a 'two-world view' – agree with proponents of an anti-metaphysical interpretation that any metaphysical interpretation of Kant's idealism is, *as such*, committed to holding that appearances and things in themselves are numerically distinct entities. Proponents of both camps assert that their respective opponents' construal of the relation between appearances and things in themselves is absurd.²⁰ The interpretation I have sketched so far suggests that we can combine an account that has ontological implications with the rejection of numerical distinctness: *one and the same thing* can be considered as having certain (spatiotemporal) properties qua object of human sensibility and and as having other properties apart from this relation to our sensibility. What one conceives as ontologically distinct here are not two distinct entities but two distinct characters or constitutions ('Beschaffenheiten' is Kant's word) that are exhibited by one and the same entity.

In what follows I will argue that this metaphysical 'two-aspect' interpretation is correct. In section 2 I will argue that we should not accept the idea that appearances and things in themselves are numerically distinct entities. In sections 3 and 4 I will show that Kant's talk about things in themselves cannot be ontologically deflated. In section 5, I will show that my interpretation is preferable over other (more ambitious) metaphysical 'two-aspect' readings.

2: The numerical identity between appearances and things in themselves

There are two contexts in which Kant straightforwardly affirms the numerical identity between appearances and things in themselves. The first one relates to his discussion of freedom. In the critical writings, Kant consistently holds that the problem of free will may be solved (only) through "the distinction... between things as objects of experience *and those same things* as things in themselves" (B XXVI; my emphasis).²¹ The second context is independent of the concern with freedom and relates to the implications that transcendental idealism has for all objects (including chairs and donkeys). If we abstract from the relation that an object has to our

knowledge is at once trivial and 'improper', for it follows from the mere *negation* of a feature that is contained in the concept of an object of human experience: namely, the feature of causal determination that applies (Kant argues) to objects solely by virtue of their membership in an objective temporal framework. To say that things in themselves lack this feature is thus to say only that things in themselves are not subject to the conditions of time as a form of human sensibility. This gives us no positive, substantive insight into noumenal reality.

²⁰ See, on the side of proponents of the 'two-world view', Van Cleve 1999, p. 8; and, on the side of proponents of the 'two-aspect view', Allison 2004, p. 16; Hudson 1994, p. 23; Korsgaard 1996a, p. 203; Matthews 1969, p. 204.

²¹ Compare CPR, A 537/B 565; A 538-9/B 566-7; A 540-1/B 568-9: "In its empirical character...this subject, as appearance, would have to conform to all the laws of causal determination. (...) In its intelligible character...this same subject must be considered to be free...from all determination through appearances". See also CPR, A 545/B 573; A 546-7/B 574-5; GMS, 4: 456 and KpV, 5: 95: "...no other way remains but to consider that the existence of a thing, so far as it is determinable in time, and therefore its causality, according to the law of natural necessity, belong to appearance, and to attribute freedom to the same being as a thing in itself". Watkins (2004, p. 318) cites a note which represents Kant as saying that freedom and causal necessity relate to *different* objects. I find it generally problematic to adduce unpublished notes as evidence for what Kant's considered views are; the textual evidence from the published writings tells decidedly against taking this particular note at face value.

faculties of sensibility, we get a concept of *that thing* as a noumenon: "The understanding, when it entitles an object in a certain relation mere phenomenon, at the same time forms, apart from that relation, a representation of an object in itself" (CPR, B 306). Moreover, the very concept of something's appearing to us implies that *this thing* must also have a character in itself:

For if the senses represent to us something merely as it appears, *this something* must also in itself be a thing. (CPR, A 249; my emphasis)

All possible speculative knowledge of reason is limited to mere objects of experience. (...) We cannot know *these objects* as things in themselves, we must yet be in position at least to think *them* as things in themselves; otherwise we should be landed in the absurd conclusion that there can be appearance without anything that appears. (CPR, B XXVI; my emphasis)

...the object as appearance is to be distinguished from itself as object in itself. (CPR, B 69)

Attempts made by proponents of the two-world view to explain away this textual evidence are, as far as I can see, unconvincing.²² This raises the question of whether the two-world view can claim textual support that is on par with that which speaks in favor of a two-aspect view.

One passage that seems to elicit the intuition that Kant is committed to the numerical distinctness of appearances and things in themselves stems from the Groundwork:

...we must admit and assume behind the appearance something else that is not an appearance, namely, the things in themselves; although we must admit that as they can never be known to us except as they affect us, we can come no nearer to them, nor can we ever know what they are in themselves. This must furnish a distinction, however crude, between a world of sense and the world of understanding, of which the former may be different according to the difference of the sensuous impressions in various observers, while the second which is its basis always remains the same. (GMS, 4: 495)

It is, presumably, the idea that the thing in itself is something 'behind' the appearances, and the appeal to a distinction between a 'world of sense' and a 'world of the understanding', that is taken to support the numerical distinctness view. But in fact the numerical distinctness view is precluded by Kant's line of thought in this passage: he says *of things in themselves* that we know *them* only as *they* affect us, i.e., insofar as *they* appear to us. Thus, he holds that the *same object* that appears to us is also, considered apart from the way in which it appears, unknowable. With the interpretation I sketched in the preceding section, we can say that what is 'behind' the appearance is the constitution ('basis') that *the object that appears* has independently of its relation to our sensibility. Kant's talk about 'two worlds' can be read as a metaphor for two perspectives that differ radically in the kind of cognitive access to objects that they afford.

Another passage, adduced by Van Cleve²³, is from the First Critique:

...we must bear in mind that the concept of appearances...justifies the division of objects into phenomena and noumena, and so of the world into a world of the senses and a world of the understanding...and indeed in such manner that the distinction does not refer merely to the logical form of our knowledge of one and the same thing, according as it is indistinct or distinct, but to the difference in the manner in which the two worlds can be first given to our knowledge, and...to the manner in which they are in themselves generically distinct from one another. (CPR, A 249)

Now, notice that this passage *immediately* continues as follows: "For if the senses represent to us

²² Van Cleve, for instance, suggests (1999, p. 145) that Kant's double-aspect language is a "rhetorical device for expressing what is actually a double-object view" and that when Kant says, 'these same objects', what he really means is "these same-sounding objects". This is purely ad hoc and entirely unconvincing.

²³ See Van Cleve 1999, p. 146.

something merely as it appears, this something must also in itself be a thing". Van Cleve suppresses this sentence in his quote. But it undermines his idea that Kant in the above passage considers things in themselves and appearances numerically different entities. Again, Kant's talk about two worlds here is to be understood in terms of two different manners of cognitive access. What Kant wants to emphasize, against the Leibnizians, is that our forms of sensibility do not give us a 'confused' access to the same objects that are 'clearly' represented through mere concepts. Rather, Kant argues, our forms of sensibility yield both (epistemic) conditions on human knowledge of objects and (ontological) conditions on objects of human knowledge: they determine a distinctive character for objects by virtue of which these objects become knowable to us. *These same objects* lack this anthropocentric character when represented by a divine intellect whose knowledge is not dependent on conditions of sensibility. On this picture, objects as viewed by us and objects as represented by a pure intellect are "generically distinct from one another": i.e., they differ with respect to the constitution (*Beschaffenheit*) that they exhibit. But this does not imply that Kant considers phenomena and noumena numerically distinct entities.²⁴

In sum, there is no conclusive textual evidence for the numerical distinctness view. By contrast, as we saw, the numerical identity view is supported unambiguously by numerous passages. Thus, attending to the details of the text strictly tells against the 'two-world' view. Notice, though, that the fact that there is a relation of identity between a thing as it appears and that same thing as it is in itself does not rule out that there are things in themselves that *are* numerically different from appearances, namely, *things that do not appear* to us:

If we entitle certain objects, as appearances, sensible entities (phenomena), then since we thus distinguish the mode in which we intuit them from the nature that belongs to them in themselves, it is implied in this distinction that we **[either: entweder]** place the latter **[eben dieselbe]**, considered in their own nature [*Beschaffenheit*], although we do not so intuit them, or **[oder]** that we place other possible things, which are not objects of our senses but are thought as objects merely through the understanding, in opposition to the former, and that in so doing we entitle them intelligible entities (noumena). (CPR, B 306)

Doubtless, indeed, there are intelligible entities corresponding to the sensible entities; there may also be intelligible entities to which our sensible faculty of intuition has no relation whatsoever.... (CPR, B 308-9)

Kant here distinguishes between two cases. In the first one (i), something that appears to our senses is correlated with a corresponding thing in itself, and this is a relation of numerical identity: if something appears, this something ('*eben dasselbe*') must have a character that is not appearance. The second case (ii) relates to a possible being (such as God) that does not appear, i.e., that has no relation to our sensibility. Such an entity, if it did exist, would not be identical to any appearance and hence would not be identical to an object that appears to us.

Now, I think that the real source of resistance toward the numerical identity interpretation derives from the worry that we cannot make good philosophical sense of this interpretation. Here a number of different concerns must be distinguished. One common worry is that if Kant did identify things in themselves and appearances, he would run into the contradiction that one and the same thing is spatial and non-spatial.²⁵ Now, Kant himself addresses this worry, and this strongly suggests that he holds the view against which this worry is directed. He says:

²⁴ Van Cleve asks (1999, p. 146): "How is one to develop a one-world view in a way that does not run afoul of Kant's opposition to Leibniz's doctrine of two modes of access?" The answer is that for Kant our only cognitive mode of access to objects is via a combination of sensible and conceptual elements that does not represent objects 'indistinctly' but allows us to represent objects, as they appear to us, distinctly in knowledge claims.

²⁵ See van Cleve 1999, p. 8 and Watkins 2004, p. 323.

The proposition, that all things are side by side in space, is valid under the limitation that these things are viewed as objects of our sensible intuition. If, now, I add the condition to the concept, and say that all things, as outer appearances, are side by side in space, the rule is valid universally and without limitation. Our exposition...establishes the reality, that is, the objective validity, of space in respect of whatever can be presented to us outwardly as object, but also at the same time the ideality of space in respect of things when they are considered in themselves through reason, that is, without regard to the constitution of our sensibility. (CPR, A 28/B 44)

...in an appearance the objects, nay even the properties that we ascribe to them, are always regarded as something actually given. Since, however, in the relation of the given object to the subject, such properties depend upon the mode of intuition of the subject, the object as appearance is to be distinguished from itself as object in itself. (CPR, B 69)

In general, there is no contradiction in supposing that one and the same object has a property that it owes to a certain relation and lacks that property independently of this relation. Kant makes use of this general idea in his idealism: he holds that there is no contradiction in saying that one and the same object has a spatiotemporal constitution as an object of, or in relation to, human intuition and lacks this constitution in itself, that is, independently of its relation to human sensibility or insofar as it is not an object of human intuition.

Proponents of the two-world view may respond that this idea renders spatiotemporal properties illusory²⁶: if spatiotemporal properties depend for their existence on an object's relation to human sensibility, these properties do not really exist at all. I find it difficult to separate this objection from the sweeping accusation that Kant's idealism *as such* is confused or blatantly unacceptable: for the objection seems to trade on the idea that spatiotemporal properties exist apart from their relation to us, which implies that space and time are not mere forms of human sensibility; and this idea strikes at the heart of Kant's idealism. In order to assess it, one would have to investigate here Kant's arguments for the ideality of space and time, a task that lies beyond the scope of this paper. But we can at least note how Kant himself would respond to the worry that his doctrine implies that spatiotemporal properties are mere illusions. He argues that space and time, when considered as mere forms of human sensibility, are the only objective, intersubjective conditions of all our (theoretical) knowledge claims; and he deems this sufficient to defend the "empirical reality" of space and time. The conviction, held by Kant's transcendental realist opponent, that unless an object has some property 'absolutely' (apart from its relation to a knowing subject) the object 'really' lacks this property begs the question against Kant's entire project.²⁷ Kant's crucial point is that a certain degree of mind-dependence *saves rather than destroys* our a priori and empirical cognition. (I return to this issue in section 4 below).

Two further, related worries about the numerical identity interpretation concern the idea that we can speak of one and the same thing 'qua' appearance and 'qua' thing in itself. The idea that one and the same thing has two different characters 'qua' x and 'qua' y is not generally unacceptable, as we can see from the case of pain: one and the same state can be considered as having a phenomenological character (*qua pain*, it is piercing) and a neurological character (*qua c-fiber stimulation*, it falls under neurological laws). But this is because once pain has been shown to be c-fiber stimulation, we can attribute the properties of pain to c-fiber stimulation and vice versa (e.g., we can say that c-fiber stimulation is piercing). The relation between appearances and things in themselves cannot be understood analogously. The spatiotemporal

²⁶ See van Cleve 1999, pp. 147-9.

²⁷ On this point I am in full agreement with Allison 2004, p. 44.

criteria that we use for determining the identity of appearances²⁸ relate to properties that we *cannot* attribute to things in themselves (since things in themselves are not in space and time). So here the idea that one and the same thing has two characters (one qua appearance, one considered in itself) seems elusive: the relation between pain and c-fiber stimulation respects whereas the relation between appearances and things in themselves violates the indiscernibility of identicals.²⁹ Moreover, not only do appearances have properties that things in themselves lack; it is also true that appearances *share no properties* with things in themselves. Hence, there is no property by appeal to which appearances and things in themselves could be identified. How, then, can we meaningfully posit a relation of identity here?³⁰

I think that Kant's response to these worries comes to light in the following passages:

...objects in themselves are quite unknown to us, and...what we call outer objects are nothing but mere representations of our sensibility, the form of which is space. The true correlate of sensibility, the things in itself, is not known, and cannot be known, through these representations... (CPR, A 30/B 45)

...the word appearance must be recognised as already indicating a relation to something, the immediate representation of which is...sensible, but which, even apart from the constitution of our sensibility...must be something in itself, that is, an object independent of sensibility. (CPR, A 250-1)

In fact, if we view the objects of the senses as mere appearances...then we admit at the very same time that a thing in itself underlies them, although we are not acquainted with this thing as it may be constituted in itself, but only with its appearance, i.e., with the way in which our senses are affected by this unknown something. (Prol, 4: 314-5)

At the same time beyond these characteristics of his own [empirical, phenomenal] subject, made up of mere appearances, he [man] must necessarily suppose something else as their basis, namely, his ego, whatever its characteristics in itself may be. (GMS, 4: 451)

These passages suggest the following picture. When we represent objects, something 'is given to us', i.e., affects our senses. The nature that this object has in itself remains unknown to us: the object that affects us, considered in itself, is "an unknown something". It is *precisely this unknown something* that we turn into an object of our knowledge by making it conform to the character of our cognitive faculties, i.e., by 'imposing' features on it that answer to the character of our representational faculties (and that it lacks in itself). In this manner, we become acquainted with the object, even though only as it appears. The very idea of an appearance directs us to something which appears and which is not itself appearance: that is, it directs us to there being a constitution that an object has in itself and that is the *basis or 'correlate' that underlies* the constitution it has as an appearance.³¹

Notice that Kant's frequent claim that 'appearances are nothing but mere representations'

²⁸ One must distinguish between (1) criteria for numerical identity *at a time* and (2) criteria for numerical identity *over time*, through change. Concerning (1), Kant says that we can distinguish objects by appeal to the difference between the spaces that they occupy (CPR, A 379/B 329), but this is not very informative unless we know how to distinguish between different spaces. When he discusses numerical identity, he is standardly concerned with (2); see CPR, A 362-3. Here it must be noticed that Kant's empirical criteria for the numerical identity of objects over time might *not* support a 'naïve' view according to which the objects of individuation are the familiar objects of ordinary parlance, such as chairs and tables: if by 'objects' we mean 'substances' in the schematized sense, Kant's official doctrine requires that the objects of individuation be *absolutely* permanent (CPR, A 285-6/B 229).

²⁹ Watkins 2004, p. 325 presses a worry along these lines against the numerical identity view.

³⁰ See Bojanowski 2006, p. 141 and Pogge 1991, pp. 495-7 for an expression of this second worry.

³¹ Note that the idea that an object's phenomenal constitution has its basis in its noumenal constitution does not, pace Robinson 1994, imply the strong idea (on occasion invoked by Findlay 1981) that there is a one-to-one mapping or isomorphism between the two characters.

does not imply that the appearance is a separate, mental entity over and above the (extra-mental) object that affects our senses.³² Such an interpretation would be strictly incompatible with Kant's commitment to the empirical reality as well as the transcendental ideality of space and time, namely, to the idea that objects are spatiotemporal in relation to human sensibility but lack spatiotemporal properties independently of their relation to human sensibility (see again CPR, A 28/B 44). This idea requires that we can attribute spatiotemporality to the very same object that is, in itself, not spatiotemporal, and this rules out the possibility that the object to which we can attribute spatiotemporality is a representation of ours: for we cannot say of a representation of ours that *it* has a character in itself that is independent from the character of our representational faculties. Thus, we must read Kant as claiming that our cognitive faculties 'impose' a certain character *on the object* that affects our senses rather than *on the representations* that this object affords us with. Passages in which he refers to appearances as 'nothing but representations' should be understood as emphasizing his idealist view that the way in which the object appears to us is determined by our representational faculties: the spatiotemporal properties of the object which provide the (sole) content for our cognitively determinate representations of objects do not have a mind-independent ontological status (see CPR, A 42/B 59; CPR, A 492/B 520; I will return to this issue in section 3 below). Appearances are (as Kant sometimes puts it) "in us" in the sense that objects (that exist and that have a mind-independent character 'outside us') appear to us only by means of our 'inner' states of consciousness; through this means, they are made to conform to the character of our 'inner' representational faculties.³³

If we understand Kant's account in this way, we can provide a response to the second of the above worries. This worry was to the effect that we cannot consider appearances and things in themselves identical since there is no shared property by appeal to which they could be identified. Kant's point is that the properties an object has qua appearance are *grounded in or based on* the properties it has in itself (or, rather, in the way the properties it has in itself affect our representational faculties). The fact that the object in itself is the correlate of the object qua appearance seems sufficient to underwrite the identity of appearance and thing in itself even if there is no property that one and the same thing has both qua appearance (in relation to human sensibility) and in itself (independently of this relation).

Now recall the first worry, which claims that Kant is not entitled to identify appearances and things in themselves because the character possessed by appearances cannot be attributed to objects considered in themselves. There is no way to avoid the concession that Kant's conception of the identity between appearances and things in themselves violates the indiscernibility of identicals: the object qua appearance is, whereas the object in itself is not, spatiotemporal. But I think that Kant would hold that this is an insurmountable problem for an identity claim only within a realist framework in which objects have one 'absolute' character independently of their relation to knowing subjects. In Kant's (transcendental) idealist framework, the way in which objects are constituted depends on the knowing subjects for which these objects *are something* (namely, a determinate object of cognition): and this means that one and the same object has or lacks a certain character depending on the type of subject that cognizes the object, i.e., depending

³² This kind of reading is most forcefully defended by Guyer 1987, p. 281, p. 291, pp. 334-5. Since Guyer bases his reading mostly on the A-edition (in particular on the Fourth Paralogism), it is worth noticing that Kant affirms *already* in the A-edition that the bearer of spatial properties is non-spatial in itself. Hence, the A-edition is incompatible with the idea that what is spatial is a representation of ours.

³³ I take this reading to be in accordance with the one suggested by Longuenesse 2001, pp. 20-1 (her footnote 9).

on those features of the subject's cognitive capacities that enable her to know and understand things. Hence, objects that have a certain character in relation to *one* type of knowing subject lack this character, and acquire a different character, in relation to a *different* type of knowing subject: in the terms of the passage quoted above, objects are "generically different" for different types of cognizers. This shows that one can undermine, at least from within Kant's idealist system, the objection to the numerical identity interpretation that derives from a reliance on the indiscernibility of identicals. Moreover, there is – as we saw – a powerful *positive* reason for identifying the object that appears with the thing in itself: the character that the object has in itself underlies or grounds the character it has qua appearance. What appears (to beings like us) in spatiotemporal form *is* the non-spatiotemporal character that the object has in itself.

My interpretation draws heavily on the idea that *an object is given* to human sensibility and that our cognitive faculties determine a specific (phenomenal) character for this object. This idea might seem problematic: one might complain that on Kant's account, it is through synthesizing given intuitions that our representations come to have intentional objects in the first place, namely, objects that have a determinate spatiotemporal constitution and position.³⁴ The possibility of there being such objects depends on the possibility of synthesizing given intuitions under the categories; objecthood is constituted and 'imposed' through the exercise of our cognitive capacities. Hence, it is illegitimate to make use of the idea that an object is given to us prior to our exercise of these capacities.

Note here that if it were indeed illegitimate to invoke a notion of objecthood that differs from the kind of objecthood which depends on the exercise of our cognitive capacities, this would not only tell against the idea that objects qua noumena are identical to objects qua phenomena; it would also rule out the idea that noumena and phenomena are numerically distinct objects. That is: both the two-aspect and the two-world view must appeal to a notion of objecthood that is independent of our cognitive faculties. The worry under consideration might thus be aimed at encouraging the idea that both the two-aspect and the two-world view are false, and that we cannot settle the question concerning the relation between appearances and things in themselves.³⁵ Now, it is surely correct to point out that we do not have any cognitively determinate handle on the kind of objecthood exhibited by things considered apart from their relation to our cognitive faculties. But this does not preclude the idea that we can consider things in themselves as objects. What *would* preclude this idea is the claim that we have no conceptual resources whatsoever for thinking of things in themselves as objects. This stronger claim, however, is indefensible, for Kant clearly thinks that we can form the indeterminate and abstract thought of an object in general ('überhaupt') without drawing on sensible intuition (see CPR, A 256/B 312; I shall discuss this point in greater detail below in section 3). Moreover, there is a compelling reason for holding that we *must* be able to think (in a cognitively indeterminate fashion) of that which affects our senses as an object: this is required to make sense of Kant's idea that our representation of objects is receptive rather than (as the representation of a divine intellect) creative. We do not bring objects into existence by intuiting them; rather, our intuiting objects depends on objects already being in existence and affecting our sensible faculties:

³⁴ For a lucid formulation of the relation between Kant's notion of appearances and the *intentionality* of human (empirical) thought, see Aquila 1979 and 1981. I am not attributing the objection I am considering here to him.

³⁵ For such a 'quietist' interpretation, see Bojanowski 2006, who dismisses the question of whether appearances and things in themselves are identical or different as meaningless.

...intuition takes place only insofar as the object is given to us. This...is only possible, to man at least, in so far as the mind is affected in a certain way. (CPR, A 19/B 33)

If we do not...treat [space and time] as objective forms of all things, the only alternative is to view them as subjective forms of our...intuition, which is termed sensible, for the very reason that it is not original, that is, is not such as can itself give us the existence of the object.... Our mode of intuition is dependent upon the existence of the object, and is therefore possible only if the subject's faculty of representation is affected by that object. (CPR, B 71-2)³⁶

It is when, and only when, our sensibility is affected by things that possess objecthood in themselves that we may cognize these things as objects – not, however, by mirroring in our representations the objecthood they possess in themselves but by determining for these (same) things a 'generic' objecthood that mirrors the generic character of our cognitive faculties.

So far, we have seen that a close examination of worries about the tenability of the numerical identity interpretation confirms the idea that this is the correct interpretation of Kant's idealism: this is because the objections to this view that I have considered so far can be defused from within Kant's idealist system. But there is a further, striking problem about the numerical identity view which cannot similarly be disarmed. The worry arises from the following simple argument: (I) We can individuate appearances. (II) Appearances and things in themselves are identical. (III) We can individuate things in themselves.

The problem here is that (III) seems to imply that we have synthetic knowledge of things in themselves, something that Kant cannot allow for given his doctrine of noumenal ignorance. The following passages from the Paralogisms clearly show that Kant rejects (III):

If...we seek to extend the concept of dualism, and take it in the transcendental sense, neither it nor the two counter-alternatives – pneumatism on the one hand, materialism on the other – would have any sort of basis, since we should then have misapplied our concepts, taking the difference in the mode of representing objects, which, as regards what they are in themselves, still remains unknown to us, as a difference in the things in themselves. Though the 'I', as represented through inner sense in time, and objects in space outside me, are specifically quite distinct appearances, they are not for that reason thought as being different things [in themselves]. Neither the transcendental objects which underlies outer appearance nor that which underlies inner intuition, is in itself either matter or a thinking being, but a ground (to us unknown) of the appearances which supply to us the empirical concept of the former as well as of the latter mode of existence. (CPR, A 379-80)

But if we consider that the two kinds of objects [the soul, as object of inner sense, and bodies as objects of outer sense] thus differ from each other, not inwardly but only insofar as one appears outwardly to the other, and that what, as thing in itself, underlies the appearance of matter, perhaps after all may not be so heterogeneous in character... (CPR, B 427-8)

Kant holds that distinct appearances may, for all that we can theoretically know, have the same underlying noumenal ground. Likewise, he must hold that the representation of a given spatiotemporal object may, for all we can theoretically know, have its ground in a number of distinct noumena affecting the mind (and hence underlying the appearance). This shows that (III) is – in Kant's system – false. Now, since Kant is committed to (I), this seems to leave us with the option of either rejecting (II) or rejecting the inference from (I) and (II) to (III). But rejecting (II) is tantamount to straightforwardly giving up on the numerical identity interpretation. Rejecting the inference from (I) and (II) to (III) requires that our individuation of appearances has no strict

³⁶ One might object (cf. Abela 2002, p. 140) that in these passages Kant is talking about the affection of the empirical mind by empirical objects. But the passages quoted above on p. 9 show that this is incorrect: Kant holds that "our senses are affected by this unknown something" which is the object in itself (Prol, 4: 314-5).

implications for the individuation of the things in themselves that are supposed to stand in a relation of identity to appearances. This suggests that noumenon A may be identical to both appearance B and appearance C, while B and C are different appearances. This is a violation of the transitivity of identity; hence (II) employs an obscure notion of identity, and so the numerical identity interpretation seems to collapse into obscurity after all.³⁷

This problem calls for a subtle response. We can begin by emphasizing that Kant does not hold that we have any reason to take seriously the idea that different appearances *are* identical in themselves, or that one and the same appearance *is* grounded in different noumena. His doctrine of noumenal ignorance, as based on the idea that our only theoretically fruitful criteria of identity depend on our cognitive faculties, only commits him to the claim that we have no theoretical grounds for either affirming or denying these possibilities. For instance, it may be the case, for all we can theoretically know, that Spinoza is right and that at the level of noumenal reality Everything is One. These are the kinds of metaphysical issues that Kant explicitly discourages us from pursuing on a theoretical basis (that is, by engaging in episodes of conceptual analysis).

Theory must remain silent on questions of noumenal identity; but Kant holds that we have *practical* criteria for fixing the identity of certain special things in themselves: namely, those beings to whom we attribute rational faculties and, specifically, a moral character. These criteria derive from our awareness of the moral law. Kant argues that we know the moral law to be valid; and since the moral law imposes valid demands on us, we have reason to consider its strict implications valid as well.³⁸ One such implication concerns the legitimacy of the idea that the addressees of moral 'oughts' – finitely rational, free agents – are worthy of praise, blame or even punishment for their free choices to respect or reject the reasons represented by moral 'oughts'.³⁹ So much is part of the very idea of moral obligation. Now, praise, blame and punishment in turn presuppose the legitimacy of identifying appearances and noumena, i.e., of holding that human beings that we distinguish from one another as bringing about observable effects in space and time *also* differ from one another qua free noumenal agents, and of holding that someone we identify empirically as acting in space and time has a persisting moral personality which is the noumenal cause of these acts. Hence, our valid idea of moral obligation provides us with grounds for identifying appearances and noumena.⁴⁰ Thus, Kant affirms the

³⁷ See Van Cleve 1999, p. 149 for the worry that the numerical identity view violates the transitivity of identity. One might wonder why I take the violation of the transitivity of identity to be damning to the numerical identity interpretation after I argued that we should swallow the violation of the indiscernibility of identicals. The answer is that the rejection of indiscernibility follows straightforwardly from the idea – which lies at the heart of Kant's idealism – that one and the same object can *be* different in relation to different knowing subjects. There is no similar motivation for the rejection of transitivity; the notion that B and C are two different appearances but one noumenon (A) seems to me to call for the verdict that we are dealing with two different sets of objects here.

³⁸ This is the point of Kant's famous idea that the moral law is the ratio cognoscendi of freedom (see, for instance, KpV, V: 4). How this idea is to be understood is the topic of the final chapter (XIII) of my dissertation.

³⁹ For the claim that the moral law legitimizes our ideas of blame, praise and punishment, see KpV, V: 37; 98-99.

⁴⁰ One could object that the fact that Kant's theory implies that numerically different empirical agents might, for all we can theoretically know, be identical qua things in themselves is manifestly intolerable (cf. Strawson 1966, pp. 248-9). But here a Kantian can respond that a focus on *theoretical*, *empirical* criteria of identity cannot answer to our *practical* intuitions concerning personhood anyway: spatiotemporal criteria of what persists through change do not allow us to identify a persisting bearer of moral character (consider the problems raised by Parfit 1986; see also footnote 28 above). We can employ the sense of identity deriving from our notion of moral personhood to correlate a noumenal being, as a source of free agency, with an appearance that we observe as causing actions in space and time. Thus, the claim that "there is no viewpoint from which we can identify something as an

following argument: (I) We can individuate appearances. (II') With regards to beings who, in themselves, are free and rational, the appearance and the thing in itself are identical. (III') We can individuate those things in themselves who are free and rational.

Since the grounds for identifying things in themselves envisaged here are based on practical rather than theoretical principles, (III') respects Kant's commitment to noumenal ignorance. This explains Kant's affirmation of the numerical identity view in those passages where he is concerned with human freedom (see p. 5 above).

But what about those passages in which the affirmation of the numerical identity view occurs independently of the concern with freedom, as an articulation of the upshot of transcendental idealism as such ("For if the senses represent to us something merely as it appears, this something must also in itself be a thing")? Kant here explicates the very concept of an appearance, as applying to *all* phenomenal objects, including those (like chairs and donkeys) that are not free and responsible in themselves and to which practical criteria of identity are inapplicable. Hence, with regard to those passages where Kant affirms the numerical identity view independently of the concern with freedom, we face the original problem. In order to avoid the consequence (which would violate Kant's noumenal ignorance thesis) that we can individuate things in themselves as such, on theoretical grounds (rather than by appeal to those practical criteria that apply only to certain special things in themselves), we are forced to choose between two options, both of which are fatal to the numerical identity interpretation: we must *either* disavow the claim (II) that things in themselves and appearances are identical *or* make room for the utterly obscure possibility that different appearances may be identical in themselves.

I think that adopting an agnostic stance about (II) might be our best bet here. With regard to the question of how appearances *as such* are related to things in themselves, Kant is not entitled to the claim that *one* object exists in itself and appears to us. But nothing hinges on this unrestricted identity thesis about the relation between phenomena and noumena. Kant can be content with the claim that for every object that appears to us in spatiotemporal form, we know that its phenomenal character is grounded in some underlying noumenal reality (which we cannot individuate on a theoretical basis). This claim is weaker than (II), for it does not by itself support a numerical identity claim between *some one* appearance and *some one* thing in itself. But it does not imply the numerical distinctness interpretation (the 'two world view'): the bare epistemic possibility, which we have no reason to take seriously, that a given appearance might be grounded in different noumena, or that different appearances might be grounded in the same noumenon, does not support the view that appearances and things in themselves *are* numerically distinct entities. Furthermore, the idea that every appearance has an underlying noumenal ground can be turned into a strict numerical identity claim with regard to those beings to whom we can attribute a rational character. As we saw, if we restrict our attention to this class of things in themselves, we can (on the basis of practical grounds, and hence without violating the noumenal ignorance thesis) endorse the notion that with regard to every being that falls into this class there is *some one* underlying noumenal ground corresponding to its spatiotemporal appearance.

It turns out, then, that the numerical identity interpretation can be affirmed only with respect to a restricted domain of objects; but it seems to me that the sense in which it can be

empirical object and the same thing as a thing in itself" (Allais 2010, p. 17) is too strong. It generalizes from a lack of *theoretical* grounds to a *complete* lack of grounds for supporting identity claims about noumena or about phenomena and noumena. See CPR, B 408-10; B 421; B 424-6.

affirmed – namely, with regard to the identity of appearances and things in themselves in the domain of free agents – is the only one that matters to Kant and that should matter to us.

3: The elusiveness of a 'merely epistemological' interpretation

On the basis of what I have argued thus far, it seems that one can refrain *both* from representing the nature of things in themselves in any even 'minimally informative' way *and* from saying that things in themselves are numerically distinct from appearances, without believing that the distinction between things in themselves and appearances lacks ontological import. In this section, I want to consider the question of whether a merely epistemological interpretation, that construes this distinction as lacking any such ontological import, is possible.

Here we must acknowledge first of all that the notion that *things in themselves exist* is ambiguous, because of the aforementioned distinction between two types of noumena. As we saw, the notion that things in themselves exist could mean: (i) Objects that appear to us also have a character in themselves. Or it could mean: (ii) There are objects (such as God) that do not appear to us. As far as (ii) is concerned, I agree that Kant's doctrine is 'merely epistemological': we cannot know (at least not on the basis of theoretical grounds) whether or not noumena that do not appear to us exist. But Kant is clearly not similarly agnostic with regard to (i). He says:

...the understanding, just by the fact that it accepts appearances, also admits to the existence of things in themselves. (Prol, 4: 315)

...the word 'appearance' must...indicate a relation to something the immediate representation of which is, to be sure, sensible, but which in itself, without this constitution of our sensibility (on which the form of our intuition is grounded) must be something, i.e., an object independent of sensibility. (CPR, A 251-2).

Thus it follows from the fact – which, Kant thinks, is established by his critical philosophy – that we know objects only as they appear that these objects also have a constitution in themselves: things in themselves exist in the sense of (i).

Now, Allison's self-proclaimed anti-metaphysical interpretation does not deny this. This interpretation holds that we know objects only 'under one aspect', namely, insofar as they appear to us. This implies that these objects must be unknowable (to us) under another aspect; and this, in turn, implies that there *is* a kind of character that objects have and that we pick out when we use the notion of a 'second aspect'. But then it seems rather unclear in what sense Allison's interpretation can be considered 'merely epistemological'. Allison's phrase that things in themselves must be regarded as "transcending the conditions of human cognition"⁴¹ seems to involve a commitment to the idea that there *is* something ("an unknown something", as Kant puts it) which goes beyond the representational resources that our cognitive faculties provide us with. Allison claims that "one can perfectly well remain within the critical philosophy with the [concept] of the thing in itself...non-metaphysically understood".⁴² But I find it hard to see how an interpretation of the concept of the thing in itself can be considered non-metaphysical if it concedes that objects have a character, or mode of being, that we cannot know. To be sure, this concession is much less determinate than some of the metaphysical interpretations of Kant's idealism that Allison wants to reject, and it does not imply that appearances and things in themselves are distinct entities. But it does involve a commitment to the idea that the objects of our experience possess ontologically real features that we cannot cognize or positively

⁴¹ Allison 2004, p. 12.

⁴² Allison 2004, p. 73.

determine. This is surely a very robust metaphysical idea.

I can think of one reply that a defender of a non-metaphysical interpretation of Kant's idealism might make here. This reply goes as follows: sure, Kant's idealism involves a commitment to the notion that objects in themselves have a character. But admitting so much involves no commitment to there being any *further* ontological feature over and beyond the character objects possess qua appearances. That is, admitting so much does not imply that *there is more to reality* than the spatiotemporal world and the spatiotemporal constitution possessed by objects of human sensibility. In other words, Kant's unknowability thesis implies merely that objects *may, for all that we can know*, have a character that differs from the spatiotemporal character we are capable of cognizing.⁴³ This is a truly anti-metaphysical interpretation.

Now, if Kant's position were merely (as this interpretation envisages) that objects may or may not have a character that differs from their phenomenal character, he would be agnostic about whether things in themselves are in space and time, or about whether space and time exist independently of our mind. But nothing could be more alien to Kant than such agnosticism:

...space does not represent any determinations that attaches to the objects themselves, and which remains even when the abstraction has been made of all the subjective conditions of intuition. (CPR, A 26/B 42)

...space is not a form inhering in things in themselves as their intrinsic property...objects in themselves are quite unknown to us... (CPR, A 30/B 45)

Time is...a purely subjective condition of our (human) intuition...and in itself, apart from the subject, is nothing. (CPR, A 35/B 61)

If we take away from our inner intuition the peculiar condition of our sensibility, the concept of time likewise vanishes; it does not inhere in the objects, but merely in the subject which intuits them. (CPR, A 27-8/B 54)

What we have meant to say is that all our intuition is nothing but the representation of appearance; that the things we intuit are not in themselves what we intuit them as being...and that if the subject, or even only the subjective constitution of the senses in general, be removed, the whole constitution and all the relations of objects in space and time, nay space and time themselves, would vanish. (CPR, A 42/B 59)

It is...not merely possible or probable, but indubitably certain, that space and time...are merely subjective conditions of all our intuition, and that in relation to these conditions all objects are therefore mere appearance, and not given us as things in themselves which exist in this manner. (CPR, A 48-9/B 66)

If we suppose an object of a non-sensible intuition to be given, we can indeed represent it through all the predicates which are implied in the presupposition that it has none of the characteristics proper to sensible intuition; that it is not extended or in space, that its duration is not a time, that no change...is to be met with in it, etc. (CPR, B 149)

But this space and this time, and with them all appearances, are not in themselves things; they are nothing but representations, and cannot exist outside our mind. (...) ...time cannot be a determination of a thing in itself. (CPR, A 492/B 520)

⁴³ Guyer 1987, p. 289 seems to suggest that Kant's idealism reduces to such a stance of epistemic modesty (only) in the B-edition of the First Critique. Allison himself wavers on this issue. At one point (2004, p. 7) he concedes that the thesis "that we cannot be sure about the matter" (the matter being: are things in themselves different from the way we represent them as being?) is too weak. One might think that this would commit him to the claim that things in themselves *are* different from the way we represent them as being, but Allison tries to repudiate this implication by denying that Kant's commitment to the non-spatiality of things in themselves "is still a move within ontology" (ibid., p. 121). But what, then, is Kant's point here? From what he says on p. 121, it seems that Allison is interpreting Kant's commitment to the notion that things in themselves are not spatial in terms of the idea that we are *not warranted* in attributing spatial predicates to objects independently of conditions of our experience. I concede that this idea carries no ontological commitments. But I find it impossible to separate this idea from the suggestion, dismissed as too weak by Allison himself, 'that we cannot be sure about the matter'.

It is also worth noting here that Kant agrees with Berkeley (CPR, A 39/B 56; B 70-1; 274-5) that the idea that space is a mind-independent entity is inherently confused. If space considered as a mind-independent entity is indeed 'ein Unding' (a 'non-entity'), it follows that objects as they are in themselves, apart from their relation to our mind, cannot have spatial properties.⁴⁴

Thus, the problem for a merely epistemological interpretation is that Kant's tenet that things in themselves are unknowable implies that the concept of the unknowable is directed toward *something* (a character) that exists and that is unknowable. The only possible strategy for denying this implication requires the idea that Kant is agnostic about whether objects that appear to us have a character that differs from the spatiotemporal character they possess qua appearances. But, as we saw, Kant is clearly not agnostic in this sense. Thus, if we take, via (i), the notion of a thing in itself to be a shorthand clause for the notion of a non-spatiotemporal constitution that objects possess independently of their relation to our sensibility, it follows that Kant is committed to the idea that objects *have* such a constitution (i.e., that things in themselves exist in the sense of (i)). His insistence that the concept of appearances implies the existence of things in themselves shows that he accepts the claim (involving, as we might say, an existential quantifier) that there *are* properties of objects that *differ* from those that we can know.⁴⁵

Now, Allison suggests that the antidote against a metaphysical interpretation derives from the insight "that the transcendental distinction between appearances and things in themselves [is to] be understood as holding between two ways of *considering* things".⁴⁶ But understanding the distinction in this way is certainly compatible with a metaphysical interpretation: to consider objects in themselves is to think of them as having an ontological character other than the spatiotemporal character that they owe to their relation to us. It is Allison who faces an important question here: what is supposed to remain, for the content of our thought, after we abstract from the relation of objects to our sensibility? My metaphysical interpretation holds that what we are left with is a cognitively indeterminate conception of an object that has a character which differs from its sensible (phenomenal) character. For there to be a meaningful contrast between a

⁴⁴ To this one might object that space and time could be construed, as on the Leibnizian view, as mind-independent relations among things in themselves (rather than as absolute Newtonian entities). But Kant thinks that in this case we would have to regard our representation of spatiotemporal features as *confused* and could not consider them a source of knowledge (CPR, A 40/B 57). I agree with Guyer, 1987, p. 360 ff. that Kant thinks that the claim that space is a *mere* form of intuition is required to save the possibility of a priori geometrical knowledge (CPR, A 48-9/B 66). If the properties examined by geometry were not based on our human conditions of spatial representation but instead pertained to things as they are in themselves, geometry could not establish necessary claims about the way objects must be (for we cannot know modal truths about mind-independent entities).

⁴⁵ Hanna 2006, p. 197 argues that Kant cannot simply infer to the existence of things in themselves by appeal to the concept of appearances, because Kant's critique of the ontological argument shows that mere concepts do not support existential commitments. Now, Kant's ontological argument clearly tells against the idea that we can infer to the existence of a type (ii) noumenon (i.e., to the existence of an object that does not appear to our senses, such as God), on the basis of analyzing its concept. But (as Allais 2010, p. 15 points out), inferring to the existence of noumena in the sense of (i), does not involve a commitment to the existence of objects that we cannot perceive. Rather, inferring to the existence of a type (ii) noumenon involves only a commitment to the existence of an unperceivable character of objects whose existence we already know synthetically, through perception. From the fact that what we perceive are mere appearances, it follows that we can analytically infer that they must also have a nature in themselves which differs from the nature they have qua appearances. *That* the objects of our awareness *are* appearances is not, of course, established by Kant on the basis of mere conceptual analysis: it is the result of the arguments advanced in the Aesthetic and the Dialectic, which seek to show that space and time are *mere* forms of human intuition.

⁴⁶ Allison 2004, p. 16. Allison here loosely follows Prauss 1974, pp. 26-8.

metaphysical and a non-metaphysical interpretation of Kant's idealism, the answer that a proponent of the latter interpretation gives to the above question must differ from mine.

Gerold Prauss, the 'founding father' of the anti-metaphysical approach, takes up this challenge. For Prauss, considering objects as appearances is synonymous with considering them, or conceiving of them, *as something*⁴⁷: to consider things as appearances is to consider them as having a certain character, but to consider them in themselves is *not* to consider them as having an alternative kind of character, for their phenomenal character is the only one that we can conceive of and that they can possess.⁴⁸ Hence, Prauss' answer to the question of what remains for our thought once the abstraction from human forms of sensibility has been made is: *nothing at all*. This gives a clear sense to the idea of a non-metaphysical construal of Kant's idealism. On Prauss' account, Kant's distinction between things in themselves and appearances has no ontological implications because the ontological ascription of a noumenal (non-spatiotemporal) character to objects is conceptually impossible.

But the problem here is, once again, that Kant himself insists that we know things only insofar as they are related to our representational faculties. This implies that objects are unknowable insofar as we abstract from their relation to our representational faculties; and this suggests that objects are a certain way, which is unknowable to us, independently of their relation to our cognitive faculties. This is the point of Kant's repeated reference to an 'unknown something' that underlies appearances. But on Prauss' approach, there *cannot be* such an unknown something; and so Prauss' analysis forces us to sacrifice a crucial feature of Kant's theory, leaving it obscure what we should make of Kant's talk about things in themselves.⁴⁹

Furthermore, the anti-metaphysical construal of the concept of the thing in itself envisaged by Prauss implies that if we abstract from the relation that objects bear to our faculties of sensibility, we lose all conceptual resources for a meaningful ascription of any kind of character.⁵⁰ There are things that Kant says which may be interpreted in this vein: he holds that concepts are devoid of content and meaning (Sinn) if they do not relate to possible objects of intuition (CPR, A 240/B 299). But we should not equate Kant's notion of an empty or meaningless concept with a contemporary notion of a concept that lacks meaning (e.g., the verificationists' dismissal of metaphysical talk as meaningless).⁵¹ When Kant says that the

⁴⁷ Prauss holds that if one considers objects in themselves, one still *considers* them, but no longer *as something*; this latter way of considering objects is supposed to be a kind of 'non-conceptual consideration' („nicht-begriffliche Betrachtung"). See Prauss 1974, p. 142-144.

⁴⁸ Prauss 1974, p. 146: "[einen anderen Charakter] koennen sie gar nicht besitzen".

⁴⁹ Prauss holds that when we consider objects in themselves, we abstract from their subject-dependency and we realize that we are 'responsible for' objects 'only among other things' (Wir erkennen an, dass wir „auch wieder nicht die einzig Beteiligten [an den Dingen] sein koennen"; 1974, p. 134; see also pp. 144-7). Something other than the cognizing subject is responsible for the way things are, but we can only acknowledge, not cognize this 'other' (ibid., p. 146). But how is talk about the subject-dependency of empirical objects, or about the 'contribution' of subjects to objects, to be understood? One possibility is that objects have certain features that are being imposed on them by our representational faculties and that they have other features apart from this imposition. This involves metaphysical notions which Prauss rejects. But without such notions, the idea of a subject-dependency of objects, or of a contribution of subjects to objects, remains utterly obscure.

⁵⁰ This is a widespread conviction. Bennett 1966, p. 24 and P. Strawson 1966, p. 16, p. 264 both claim that Kant is committed to holding that talk about noumena violates a 'principle of significance'. Melnick 1973, p. 152 holds that the concept of a thing in itself is incomprehensible. Even Walker 1978, p. 14 says that in Kant, conditions for the possibility of experience are identical to conditions for intelligible thought (but see footnote 51 below).

⁵¹ See Walker 1978, p. 131 for a very clear exposition of this point.

categories do not have meaning or content without being related to sensible intuition, his point is that they cannot give us any *determinate* object (CPR, A 247-8; B 305-6). Thoughts without intuition are empty in that they lack directedness toward a determinate *something*. But Kant accepts that the unschematized categories (i.e., the categories thought in abstraction from their relation to our sensible intuition) do have a kind of formal content, what he calls a 'form of thought' (CPR, B 306; A 254/B 310; A 147/B 186-7; A 243/B 301). Although this content fails to represent a determinate object, it does give us the bare thought of "something in general" (CPR, A 252-3) which allows us to think things in themselves as "an indeterminate something" (CPR, B 150; A 256/B 312). This thought involves no positive delineation of what objects, considered apart from our forms of sensibility, are like: it involves only what Kant calls the *negative* concept of a noumenon, i.e., the concept of something that is *not* an object of our intuition (CPR, B 306). "What our understanding acquires through this concept of a noumenon, is a negative extension" (CPR, A 256/B 312). This negative extension suffices for the indeterminate ontological idea that objects have a constitution that differs from their spatiotemporal character.

So I do not think that Prauss is right in holding that it is conceptually impossible for us to consider objects 'as something' apart from their relation to our sensibility. Now, Allison's interpretation is, in this respect, much more permissive than Prauss': Allison concedes that we can "distinguish the character that...things reveal as appearing...from the character that the same things are thought to possess when they are considered as they are in themselves"⁵². But the notion of a 'character that objects are thought to possess when they are considered as they are in themselves', and that is to be *distinguished* from the character that these objects possess qua appearances, surely seems metaphysically loaded. This is why Prauss so adamantly insists that we cannot form any thought of objects possessing a noumenal character that differs from their phenomenal character. I think that Prauss deserves credit for recognizing how radical the idea of a genuinely anti-metaphysical interpretation of the concept of the thing in itself must be, that is, for how little it allows in terms of the conceptual resources for considering objects in themselves. I agree with Allison against Prauss that Kant allows for more conceptual resources than Prauss is willing to acknowledge; we can consider things in themselves 'as (an indeterminate) something'. But I agree with Prauss against Allison that the concession that Kant acknowledges these conceptual resources implies a metaphysical interpretation of the concept of things in themselves: i.e, it implies that Kant accepts that the negative concept of a noumenon refers us, indeterminately, to an non-spatiotemporal character that objects possess in themselves.

Allison might respond that the *mere thought* that objects considered in themselves possess a character does not imply a commitment to the idea that objects *have* a character that differs from their spatiotemporal constitution. But this response must, once more, rely on the suggestion that Kant is agnostic about whether spatiotemporal properties could exist apart from the human mind, and hence about whether objects in themselves differ from the way in which they appear to us. As we saw, such agnosticism is completely alien to Kant: he affirms the mind-dependence of space and time and thereby the non-spatiality of things in themselves. Hence, Kant's notion of things in themselves cannot be understood in non-metaphysical terms.⁵³

⁵² Allison 2004, p. 56 (see also p. 17).

⁵³ A further strategy for ontologically deflating Kant's conception of things in themselves, which I find (in different forms) in Bird 1962, Grier 2001, Hanna 2006, McDowell 1998a, and Senderowicz 2005, rests on the claim that Kant uses the concept of things in themselves as a mere limiting concept which functions in ways that are similar to the regulative ideas of reason: we cannot help employing it, but we cannot know whether it is instantiated.

4: The issue of ontological priority

I now want to address a worry concerning the implications of a metaphysical reading of Kant's idealism. Karl Ameriks, a proponent of such a reading, demands that the noumenal constitution of objects has "greater ontological status"⁵⁴ than their phenomenal constitution. Allison worries that this implies that objects *really* are non-spatial, i.e., that they *merely seem* to us to be spatial.⁵⁵ The question is whether a metaphysical interpretation is committed to this implication.

I agree with Allison that we cannot saddle Kant with the idea that spatiotemporal properties are somehow illusory⁵⁶, or that our knowledge of spatiotemporal properties is epistemically inferior to the cognitive perspective of a divine intellect.⁵⁷ But I do not think that it follows from the metaphysical idea that spatiotemporal properties are mind-dependent that the mind-independent, noumenal constitution of objects has a weightier (ontological or epistemic) status than their spatiotemporal constitution. For Kant, our forms of sensibility are genuine *sources* of knowledge⁵⁸; and they provide us with a framework within which we can objectively represent a unified course of natural alterations. The concession that the spatiotemporal features of objects of our experience do not belong to objects independently of our sensible faculties does not demote the status of these features: it would do so *only if* we could supply a competing framework for objective empirical knowledge. But we have no such competing framework:

With the sole exception of space there is no subjective representation, referring to something outer, which could be entitled [at once] objective [and] a priori. For there is no other subjective representation from which we can derive a priori synthetic propositions, as we can from intuition in space. (CPR, A 28/B 44)
...the concept of alteration, and...of motion...is possible only through and in the representation of time; and that if this representation were not an a priori (inner) intuition, no concept, no matter what it might be, could render comprehensible the possibility of an alteration.... Only in time can two contradictorily opposed predicates meet in one and the same object, namely, one after the other. (CPR, B 48-9)

Moreover, Kant thinks that if there were no dependency of (the character of) objects of our

This approach conflates Kant's negative and positive concept of noumena. I discuss this distinction in detail in section 5 below. As we will see, Kant accepts that we cannot know that the *positive* concept of noumena is instantiated, but (as I argued in this section) he also holds that we know that the *negative* concept of noumena is instantiated: he holds that we know that objects have a character which differs from the features that objects owe to their relation to human intuition. Now, I agree that the positive concept of a noumenon is somewhat connected to the regulative ideal of reason: it refers us to the perspective of a divine intellect, and the ideal of reason refers us to the existence of a highest being (CPR, A 641-2/B 670-1). But again, while Kant holds that we cannot know (on theoretical grounds) whether the concept of a supersensible divine being is instantiated, he also affirms without any ontological scruples that we can theoretically know that noumena in the negative sense exist: objects that appear to us *have* a constitution in themselves. See Allais 2010, pp. 15-6 for a clear exposition of this point.

⁵⁴ Ameriks 1992, p. 334.

⁵⁵ See Allison 2004, p. 46; p. 121.

⁵⁶ See CPR, B 69-70: "Thus when I maintain that the quality of space and of time, in conformity with which, as a condition of their existence, I posit both bodies and my own soul, lies in my mode of intuition and not in those objects themselves, I am not saying that bodies merely seem to be outside me, or that my soul only seems to be given in my self-consciousness. It would be my own fault, if out of that which I ought to reckon as appearance, I made mere illusion." The conflation that Kant rejects here can be witnessed in Strawson 1966, p. 238.

⁵⁷ Ameriks 2000, pp. 255-6 claims that God's perspective on objects would be 'more adequate' than ours, and this is a claim that Kant (as far as I can see) never endorses or encourages.

⁵⁸ CPR, A 28-9/B 55: "Time and space are, therefore, two sources of knowledge, from which bodies of a priori synthetic knowledge can be derived. (Pure mathematics is a brilliant example of such knowledge...)."

experience on the character of our cognitive faculties, we would be thrown into skepticism: we would have no way of ensuring whether objects have to be a certain way (*a priori knowledge* would be impossible; see CPR, BXVII), nor even whether a given representation in fact corresponds to the nature of the object of representation (*empirical knowledge* would be impossible; see CPR, A 378-9). For Kant, his idealism, far from demoting the status of spatiotemporal objects, is the only way of saving the idea that our representations of spatiotemporal objects can be considered as yielding (*a priori* and empirical) cognition.

Those proponents of a metaphysical interpretation (like Ameriks) who understand Kant's idealism as claiming that the 'real' nature of objects is unknowable seem to share a presupposition with those (like Allison) who advance an anti-metaphysical interpretation of Kant's idealism because they worry that a metaphysical interpretation would imply that we cannot know the 'real' nature of things⁵⁹: namely, the presupposition that the concept of 'reality' has a clear, unambiguous significance. Kant does not accept this presupposition. He argues that this concept can be understood in different ways, and that we have a concept of 'empirical reality' (CPR, A 28/B 44) that does justice to our intuitive sense of mind-independence in that it allows for the applicability of intersubjective, objective criteria for truth.⁶⁰ This is compatible with there being a different notion of reality (what Kant calls "absolute reality"; CPR, A 35/B 52) that pertains to properties of objects that are independent of our sensibility. Now, the term 'absolute' might seem to imply a claim to ontological priority, but Kant himself clearly wants to avoid the idea that what is absolutely real must be *more* real (or, rather, 'more really real'⁶¹) than what is empirically real. He deems the fact that space and time are 'only' empirically real (i.e., mind-dependent), rather than absolutely real (i.e., mind-independent), *altogether irrelevant* for the epistemic status of our empirical judgments about the spatiotemporal world of sense:

This [empirical, rather than absolute] reality of space and time leaves...the certainty of empirical knowledge unaffected, for we are equally sure of it, whether these forms [of sensibility] necessarily inhere in things in themselves or only in our intuition of them. (CPR, A 39/B 56)

We can conjecture that Kant's point here is this: *all we could ever want*, in our quest for theoretical knowledge, is the availability of securely grounded criteria for objectivity,

⁵⁹ See Allison 2004, p. 121; compare Kitcher 1999, p. 431, who claims that an ontological interpretation of Kant's idealism fails because it implies that the properties imposed on objects by our faculties are "fake".

⁶⁰ Notice that Kant's idea that spatiotemporal properties are mind-dependent does not turn him into a phenomenalist or a Berkeleyan idealist, i.e., someone who holds that spatiotemporal properties depend for their existence on being actually or counterfactually perceived. Rather, the possibility of an object of appearance depends on its being an object *of a possible experience*, where this notion is explicitly broader than the notion of an object of possible perception (cf. CPR, A 522/B 550). Even if a spatiotemporal object or event could never be perceived by any human subject because our sense organs are too weak, it would still exist insofar as it could be lawfully connected to other appearances in one unified spatiotemporal framework or course of experience. For Kant the limits of perception do not determine limits of (phenomenal) existence (see Abela 2002, p. 24, pp. 237-8).

⁶¹ Walker 1983, p. 161 claims that "Kant holds that the really real world is not spatial or temporal". His idea that Kant endorses a concept of 'real reality' is closely tied to his view that "virtually all of our everyday beliefs must be in Kant's opinion false" because there is a "more ultimate" standpoint from which we recognize the mind-dependence of the character exhibited by appearances. But this betrays a rather grave misunderstanding of Kant's position: Kant clearly wants to say (CPR, A 28/B 44) that even from the standpoint from which we recognize the mind-dependence of space and time we can deem the beliefs that we hold about the spatiotemporal character of objects from the "human standpoint" *true and correct*, because we acknowledge that the intentional objects of those beliefs are (spatiotemporal) appearances rather than (non-spatiotemporal) things in themselves.

intersubjectivity, truth, or justification. Space and time uniquely (in cooperation with our faculty of concepts, the understanding) *do* make such criteria available to our human quest for knowledge of objects. Hence, the ontological status of space and time is irrelevant for someone who wants to know things about the natural world: the fact that properties such as size, shape, duration, but also substancehood and causal necessity, depend on features of our mind in no way impugns the notion that these are real features of the objects of our knowledge.

I am not saying that this idea is altogether easy to swallow. What is essential for the purposes of my overall argument is that Kant certainly rejects the objection that a metaphysical construal of his idealism, which holds that space and time depend on the human mind, implies that space and time are real only in a degenerate sense that smacks of illusion.

Now, those (like Ameriks) who claim that spatiotemporal appearances are less real than noumena might point to passages where Kant *does* seem to suggest that our empirical selves are subordinate to our noumenal selves. One such passage stems from the Groundwork:

...it is not because it [the moral law] interests us that it has validity for us (for that would be heteronomy and dependence of practical reason on sensibility, namely, on a feeling as its principle, in which case it could never give moral laws), but that it interests us because it is valid for us as men, inasmuch as it had its source in our will as intelligences, in other words, in our proper self, and what belongs to mere appearance is necessarily subordinated by reason to the nature of the thing in itself. (GMS, 4: 457-8.)

Here the claim is that practical reason deems mere practical rules, that are grounded upon contingent and merely subjective 'feelings' (of pleasure), inferior to intersubjective laws that reason prescribes independently of contingent facts about our psychology. This is a normative subordination of desires that happen to affect our empirical self to the necessary demands of morality. One cannot infer from this that Kant also envisages an ontological subordination of the intersubjective spatiotemporal character that objects necessarily have qua appearances to the character that objects have independently of their relation to our forms of sensibility. After all, Kant *also* draws the contrast between necessary, intersubjective structures and merely contingent, subjective 'seemings' within the domain of appearances, or sensibility:

Space...as condition of outer objects necessarily belongs to their appearance or intuition. Taste and colors are not necessary conditions under which alone objects can be for us objects of the senses. They are connected with the appearances only as effects accidentally added by the particular constitution of the sense organs. Accordingly, they are not a priori representations, but are grounded in sensation, and, indeed, in the case of taste, even upon feeling (pleasure and pain), as an effect of sensation. (CPR, A 29)

Measured against the standard of intersubjectivity and necessity, *both* one's contingent desire for chocolate (based on the feeling of pleasure evoked by the taste of chocolate) *and* one's contingent sensation of sweetness, *but not space and time*, (dis)qualify as mere appearances.⁶²

Thus, a metaphysical interpretation of Kant's idealism can make perfect sense of Kant's claim that space and time are (the only) empirically real sources of objective human knowledge.

5: Kantian humility: why we have no positive theoretical grasp of noumena

My interpretation of Kant's idealism is open to attack from two different angles. In sections 3 and 4, I have defended my account against the worry that it is *too metaphysical*. In this section I want

⁶² Compare CPR, B 45. It is not clear whether Kant's relegation of *colors* to mere appearances (optical illusions) persists in the *Critique of Judgment*. See Ginsborg 1998 for a perspicuous discussion of this important issue.

to address the objection that my interpretation is *not metaphysical enough*. The objection is that if I concede that Kant is committed to the existence of things in themselves (in sense (i), as explained on p. 15), I cannot confine myself to the humble idea that objects that appear to us have *some* non-spatiotemporal character: I must concede that noumenal reality consists of substances bearing non-relational properties. The textual basis for this objection stems from the section entitled Amphiboly of Concepts of Reflection. There Kant says that "as object of pure understanding...every substance must have inner determinations...which pertain to its inner reality" (CPR, A 266/B 322). This is often taken to suggest that Kant commits himself to a claim about what objects are like in themselves, considered apart from the character that is imposed on them by our *sensible* faculties (i.e., considered only as objects *of the understanding*). But is Kant here really putting forward a positive view about what things in themselves are like?⁶³

To answer this question, we must first take into account Kant's distinction between the positive and the negative concept of a noumenon. The positive concept of a noumenon is the concept of an object of a (possible) divine, non-discursive intellect. Such an intellect would intuit objects, both those that appear and those that do not appear to our senses⁶⁴, 'intellectually', apart from any conditions of sensibility and hence as they are in themselves (CPR, B 306); by contrast, cognizers that have a discursive intellect depend for cognition on receiving sensible 'data' that must be combined under the guidance of conceptual rules for object-directed thought.

We already encountered the negative concept of a noumenon in section 3: this is the concept that applies to any object that appears to us and through which we ascribe an indeterminate character that is *not* spatiotemporal appearance to this object.

Now, as we saw in sections 2 and 3, Kant's commitment to the existence of things in themselves derives directly from his commitment to appearances: if we represent objects only as they appear to us, it follows that objects must also have a constitution in themselves which is not appearance. Kant clearly thinks that this inference from appearances to things in themselves is valid *only if* one uses the negative concept of a noumenon:

The doctrine of sensibility is likewise the doctrine of the noumenon in the negative sense, that is, of things which the understanding must think without this reference to our mode of intuition, therefore not merely as appearances but as things in themselves. (CPR, B 306)
[what] we entitle 'noumenon' must be understood as being such only in a negative sense. (CPR, B 309)

Hence, Kant's commitment to the existence of things in themselves involves only a commitment to the 'negative' claim that objects that appear to us have a certain constitution that differs from

⁶³ Notice that one cannot simply hold that such an interpretation fails because of Kant's commitment to noumenal ignorance. Langton (cf. her 1998, p. 50) would respond that that her reading respects this commitment, for it implies nothing about *what* the 'inner determinations' that she ascribes to noumenal substances are.

⁶⁴ This is a controversial point. Willaschek 1998, p. 350 holds that the positive concept of a noumenon applies only to objects that do *not* appear to us; similarly, Allais 2010, p. 10 claims that "the notion of a noumenon in the positive sense is the notion of a special kind of object, distinct from the objects of which we have knowledge..., which would be an object for a different kind of intuition than ours". This seems false: Kant uses the notion 'Verstandeswesen' (which Allais renders as 'intelligibilia' and which she equates with both noumena in the positive sense and noumena that do not appear to us) both in reference to objects that appear and in reference to objects that do not appear to us (CPR, B 306, B 309; Prolegomena, 4: 315). He claims (GMS, 4: 451) that we *must* count ourselves as members of the 'intelligible world' (which underlies sensible appearances), and his argument here consists in the familiar inference from the fact that we appear to the idea that we must have a constitution in ourselves. So the notion of a noumenon in the positive sense *need not* relate to supersensible objects that are numerically distinct from those that appear to us. (See the following footnote for further discussion.)

the constitution that they owe to our 'mode of intuition'.⁶⁵ Now, those who think that we can get from this 'negative' commitment to the positive claim that objects in themselves are substances bearing intrinsic (non-relational) properties must hold that the representation of objects that we get when we abstract from our 'mode of intuition' entitles us to judge that objects in themselves (independently of our intuition) are substances bearing intrinsic properties.

I do not deny that if we abstract from our mode of intuition, and hence use the purely conceptual resources of our understanding, we get the representation of a substance as a bearer of intrinsic properties. This is undeniably the point of Kant's idea that "as subject of the pure understanding, every substance has inner determinations". We saw that the negative concept of a noumenon affords us with the representation of an 'indeterminate something'; we can now see that the representation of this indeterminate something involves the representation of a substance that has intrinsic properties. This might seem puzzling: the representation of a substance bearing intrinsic properties seems to go decisively beyond the representation of an indeterminate something. But the link between these two representations should not come as a surprise: as we saw in section 3, the negative concept of a noumenon is formed through analyzing the pure, unschematized categories, among them the pure concept of substance. Analyzing this concept affords us with the indeterminate representation of an object in general (CPR, A 284/B 340).⁶⁶

What I deny is that this feat of conceptual analysis can tell us anything about what objects in themselves *are like*, even in the minimal, abstract sense intended by the claim that objects in themselves are substances bearing intrinsic properties. Here I agree with Allison⁶⁷, although not with his reasons. For Allison, the reason that analyzing our concept of an object in general cannot

⁶⁵ More strictly, he is committed to the claim that objects that appear to us have a constitution that is independent of any conditions of sensibility, namely, an 'intelligible' constitution. See CPR, B 309: "doubtless, indeed, there are intelligible entities corresponding to the sensible entities". But this raises an interpretive problem. We saw that intelligible entities are objects of a divine intuition and as such fall under the positive concept of a noumenon. Kant holds that the positive concept of a noumenon is "problematic": we cannot know whether it is instantiated (CPR, A 254/B 310). But the claim that there 'doubtlessly *are*' intelligible entities seems to involve a commitment to the claim that the positive concept of noumena *is* instantiated. I propose the following solution. It is crucial to see that Kant defines both the concept of a sensible object (a phenomenon) and the concept of an intelligible object (a noumenon, in the positive sense) *by reference to the type of subject for which such objects could become cognitions*. So the positive concept of a noumenon implies not merely the existence of an intelligible constitution but also the existence of the kind of divine intelligence for which this constitution *would be something, namely, an object of cognition*: the positive concept of a noumenon is that "of a thing which is not to be thought as object of the senses but as a thing in itself, solely through a pure understanding" (CPR, A 254/B 310), or "of an object for a quite different intuition and a quite different understanding than ours" (CPR, A 287/B 344). Since we cannot have experience of a divine intellect, we cannot know whether a divine intellect exists; thus, we cannot know whether the positive concept of a noumenon, *as defined by reference to such a being*, is instantiated. That is: if 'intelligible object' is understood as 'object for a divine intuition', we cannot know whether any intelligible objects exist. On this basis, we might be able to understand the problematic claim at B 309 as follows: doubtless indeed, things that appear to us also have a character in themselves which is independent of any conditions of sensibility *and that would be an intelligible object of cognition for a divine intuition, if such a divine being were to exist*.

⁶⁶ That the representation achieved through the pure categories is (no more than) that of an 'object in general', or of an 'indeterminate something' (which, as we saw, provides a 'negative extension' for our understanding), is also clear from CPR, A 400: "But if I know something as simple in concept only [this is the concept of substance]...I have really no knowledge whatsoever of the object, but only of the concept which I make for myself of a something in general that does not allow of being intuited."

⁶⁷ See Allison 2004, p. 10, p. 17, p. 56; here I disagree with Allais 2010, p. 15.

establish that objects in themselves are substances bearing intrinsic properties is that *analytic* judgments cannot establish anything about things in themselves.⁶⁸ I do not see why the mere fact that a judgment is analytic should disqualify it from informing us (minimally, or trivially) about the character of objects in themselves. Rather, I think that it is the *nature of the concepts* employed in our analytic judgments about 'objects in general' that prevents these judgments from informing us even minimally about the nature of noumena. My argument is as follows: any feature that pertains to the character that objects possess in themselves must figure in the content of those representations of objects that would be achieved by a divine intellect that is capable of representing things independently of any conditions of sensibility. But the notion of a substance bearing intrinsic properties, or the notion of an ultimate subject of predication, cannot figure in the representations of a divine intellect, for a divine intellect would not represent objects according to the concepts of our understanding: it would not represent things as 'substances'.

The point here is not simply that a divine intellect would not represent objects in terms of the *schematized* concept of substance, which involves a reference to human sensibility, but that a divine intellect would not represent objects in terms of the *unschematized* concept of substance either. This is because the pure, unschematized categories are rules (or 'functions') for unifying given, sensible intuitions in general: i.e., rules that apply to any discursive intellect, regardless of the mode (e.g., space and time) in which sensible intuitions are given to it:

The same function which gives unity to the various representations in a judgment also gives unity to the mere synthesis of various representations in an intuition; and this unity...we entitle the pure concept of the understanding. (CPR, A 79/B 105-6)

The pure concepts of understanding are free from this limitation [of space and time], and extend to objects of intuition in general, be the intuition like or unlike ours, if only it be sensible and not intellectual. (CPR, B 148)

The pure concepts of the understanding relate, through the mere understanding, to objects of intuition in general, whether that intuition be our own or any other, provided only it be sensible. (CPR, B 150)

These attributes are nothing but pure categories, by which I do not think a determinate object but only the unity of the representations – in order to determine an object for them. (CPR, A 399)

A non-discursive intellect – capable of non-sensible, intellectual intuition – does not depend for cognition on given intuitions; and hence it has no need for the pure categories qua rules for unifying given intuitive representations:

This unity of intuition always includes in itself a synthesis of the manifold given for an intuition, and so already contains the relation of this manifold to the unity of apperception. And in this deduction, since the categories have their source in the understanding alone, independently of sensibility, I must abstract from the mode in which the manifold for an empirical intuition is given, and must direct attention solely to the unity which, in terms of the category, and by means of the understanding, enters into the intuition. (...) But...there is one feature from which I could not abstract, the feature, namely, that the manifold to be intuited must be given prior to the synthesis of understanding, and independently of it. How this takes place, remains here undetermined. For were I to think an understanding which is itself intuitive (as, for example, a divine understanding which should not represent to itself given objects...), the categories would have no meaning whatsoever in respect of such a mode of knowledge. They are merely rules for an understanding whose whole power consists in thought, consists, that is, in the act whereby it brings the synthesis of a manifold, given to it from elsewhere in intuition, to the unity of apperception – a faculty, therefore, which by itself knows nothing whatsoever, but merely combines and arranges the material of knowledge, that is, the intuition, which must be given to it by the object. (CPR, B 145)

⁶⁸ Allison 2004, p. 17.

We must not confuse Kant's claim that the categories "have no meaning whatsoever" in respect to the cognition exhibited by a divine intellect with the claim (which we reviewed earlier) that *for us* the categories have no meaning when they are thought in abstraction from our forms of sensibility. The latter claim, I argued, is only meant to show that we cannot, through the unschematized categories and hence independently of sensibility, represent a determinate object. But the former claim is more radical: Kant here wants to say that in relation to objects represented by a divine intellect, the unschematized categories are literally inapplicable:

...suppose an object of a non-sensible intuition to be given...to such a something not a single one of all the categories could be applied. We could not, for instance, apply to it the concept of substance, meaning something which can exist as subject and never as mere predicate. (CPR, B 149)

...there are intelligible entities corresponding to the sensible entities...but our concepts of understanding, being mere forms of thought for our sensible intuition, could not in the least apply to them. (CPR, B 309)

Thinking in terms of the unschematized categories floats free of the conditions of human sensibility *but not of conditions of sensibility in general*, i.e., of conditions of thought for a discursive intellect in general. Since the unschematized categories are rules for unifying given intuitions in general, an understanding that does not synthesize given intuitions has no use for these concepts. Since the pure concept of substance is *only and essentially* a 'function of synthesis' (CPR, A 356), a non-synthesizing intellect would not represent objects as substances.

Another way of putting this point is that in Kant's idea that "as subject of the pure understanding, every substance has inner determinations", the term 'the pure understanding' refers to a discursive understanding, which is incapable of positively representing the way objects are in themselves. This term does not refer to a divine understanding for which things in themselves would be objects of knowledge.⁶⁹ We can abstract completely from the conditions of *human* sensibility and thereby arrive at the notion of a substance that has intrinsic properties.⁷⁰ But this feat of abstraction does not support an inference to claims about what objects are like when represented by a being whose cognitive access to objects is mediated *neither* by any forms of sensibility *nor* by concepts whose sole function consists in unifying sensible intuitions.

One may object that Kant himself holds that purely relational determination *must* be grounded in something non-relational or 'inner'. Kant indeed thinks that there is a 'must' here, but it is solely the 'must' of analytic necessity: our concept of an object in general constrains us to think that relational determinations must be grounded in purely inner determinations. This analytic judgment has no ontological implications concerning the mode of being exhibited by things in themselves. To repeat, my argument is not that it is the *analyticity* of this judgment that precludes any such ontological implications. Rather, my reason for denying that the conceptual analysis of our concept of an object in general could have ontological implications is that our representation of an object in general involves the pure categories, and the way objects are in themselves would not be represented (by a divine intellect) according to the categories.⁷¹

⁶⁹ This is clear from the early sections of the *Analytic*: Kant states that *our* understanding owes the title of a *pure* understanding to the fact that it contains pure concepts for the synthesis of a manifold of intuition (CPR, A 80/B 106); and accordingly, he moves immediately from the notion that our understanding (unlike a divine understanding) is *not* a capacity for intuition to the explication of the pure categories qua functions of the unity judgment (CPR, A 67-8/B 92-3).

⁷⁰ See CPR, A 283-5/B 339-41.

⁷¹ This is why Kant says (CPR, A 284-5/B 340-1) that the necessity involved in the idea that substances have intrinsic properties is "grounded solely on abstraction". The passage continues (my emphases): "According to

These considerations allow us to expose the true scope of Kantian humility. It is clear that the anthropocentric conditions of our knowledge (namely, our forms of sensibility and the schematized categories) have no application to noumena. What is often overlooked is that the non-anthropocentric conditions that constrain our abstract concept of an object in general are *also* inapplicable to noumena. These latter conditions are the unschematized categories that would be shared by any discursive intellect, that is, by any thinking being that depends for the cognition of objects on unifying given sensible data, regardless of whether its forms of sensibility are human (spatiotemporal) or not (cf. Kant's reference to "other thinking beings" at CPR, A 27/B 43). The pure categories represent an (indeterminate) object of sensible (not necessarily human) intuition in general, and hence they would *not* be shared by an intellect whose cognition of objects is independent of any conditions of sensibility. Since the character of things in themselves is independent of any conditions of sensibility, it follows that the pure categories do not apply to things in themselves. This is to say: objects, intuited as they are in themselves by a non-discursive mind, do not fall under concepts that are *essentially* concepts for a discursive mind. The positive concept of a noumenon (which, as we saw, is the concept of an object for a divine intellect) signifies "an object for a quite different intuition *and a quite different understanding* from ours" (CPR, A 287/B 344; my emphasis); and correspondingly, the nature of things in themselves is "independent of the conditions both of our senses *and understanding* (Prol, 4: 322; my emphasis). An analysis of the pure categories, such as the notion of substance, cannot afford us with any positive determination of non-sensible reality.⁷²

mere concepts the inner is the substratum of all relational or outer determinations. If, therefore, I abstract from all conditions of intuition and confine myself to the concept of an object as such...there must still be left a concept of something which signifies no relation, but inner determinations only. **From this it seems to follow** that in whatever is a thing (substance) there is something which is absolutely inward and precedes all outer determinations, inasmuch as it is what first makes them possible.... (...) These contentions **would be** entirely justified, if beyond the concept of an object as such there *were* no further conditions under which alone objects of outer intuition can be given to us those from which the pure concept has made abstraction". This clearly shows that Kant deems the move from purely conceptual abstraction to ontological principles fallacious – on his view the move is only *seemingly* valid. However, a problem here is that Kant himself argues in the Aesthetic (CPR, B 66-7) that because our forms of sensibility give us only relational properties, we cannot cognize the inner determinations that belong to the thing in itself. Here I want to make three points. First, Kant's remark in the Aesthetic does not occupy a central role in his argument: it comes after his official arguments for the ideality of space and time, and figures only as a "confirmation" of his theory. Second, if there were a strict conflict between this passage and the treatment of noumena that comes later in the First Critique, we should give priority to that later discussion, because it is only here that Kant states his official doctrine of noumena. Third, the conflict between the two passages can be minimized by conjecturing that the dialectical point of the "confirmation" Kant adduces in the Aesthetic might be to offer a reason for the ideality of space that is accessible to the Leibnizeans: since *they* think of things in themselves as substances that are bearers of inner determinations, they ought to accept that space and time, *qua* sources of relational features, are transcendently ideal. This, together with the rejection of the Leibnizian doctrine that sensible representations are confused (CPR, A 44/B 62), might be intended by Kant as a first response to the Leibnizian doctrine, which is followed by a radicalized response in the Amphiboly. This second response embodies the full force of Kant's critique because Kant now can draw on the results of *both the Aesthetic and the Analytic* to show that Leibnizian philosophy misunderstood not only human *sensibility* but also the source and scope of our pure *concepts*.

⁷² Those, like Langton or Ameriks (2000, pp. 66-7), who claim that noumenal entities (as represented by the pure, unschematized category of substance) are the only *ontologically real* substances, do not appreciate how radically the critical philosophy breaks with commitments Kant held in his earlier writings. (This goes also for Watkins 2004, p. 354; while he does not deny that phenomenal substances are real, he holds that Kant accepts noumenal substances as "ultimate subjects.") For the critical Kant, not only our forms of sensibility but also our pure

Hence, we can now see that the negative concept of a noumenon, which affords us with the representation of an indeterminate something, is not any less 'negative' or indeterminate for the fact that it involves the thought of a substance bearing intrinsic properties. For the reasons given, this thought does not have any however minimal or abstract implications concerning the character of things in themselves. We are wholly incapable of determining any such implications through exercising our faculty of understanding. This is tantamount to saying that the concepts of our understanding fail to give us a positive concept of a noumenon (CPR, A 255-6/B 311).

This is not Kant's last word. Our lack of a positive conception of noumena is emphasized by him in his discussion of the epistemic limitations inherent in our faculty of theoretical cognition, *the understanding*: the pure concepts of the understanding are completely incapable of informing us, even minimally, about the constitution of objects that we cannot perceive. But the faculty of understanding is not our only source of concepts. With the *ideas of pure reason* – such as the notion of freedom, i.e., of an unconditioned condition of action – we *can* consider noumena in a 'positive' manner. This is because the ideas of reason are *not* functions of the unity of a given intuition: they relate to notions – the unconditioned, the infinite – that cannot possibly be met with in a sensible, discursive experience (cf. CPR, A 510/B 538) and therefore can be attributed to the conceptual repertoire of a non-discursive intellect.⁷³ Now, since these ideas cannot be instantiated in our experience, they cannot give us theoretical knowledge. But employing them enables us to consider things in themselves in a manner that is fruitful from a *standpoint* other than that of theoretical cognition, which considers us as free, responsible agents.

In the following chapter, I will show how the interpretation of Kant's transcendental idealism I have developed in this chapter bears on, and is confirmed by, an interpretation of Kant's doctrine of transcendental freedom.

concepts of the understanding are *essentially* conditions of objective discursive thought that works on the material provided by our receptive faculty of sensibility: the unschematized concept of substance is in its essence a 'function' of the unity of given sensible intuitions. Hence, objects would not be intuited *as substances* by an intuitive intellect which is unbound by conditions of receptivity; and thus, *only* phenomenal substances – albeit not necessarily substances represented through a *human* mode of sensibility – are ontologically real.

⁷³ Kant does not mean that through the ideas of reason we can arrive at *precisely* the representation of objects that can also be attributed to a divine intellect. His view is that if our discursive concept of an object in general, as represented through the pure categories, is supplemented with the ideas of reason, we can arrive at a positive conception of noumena that *contains an element* of representation that would be shared by God. See KpV, 5: 136: the idea of freedom gives an object (a positive determination) to the unschematized category of causality, namely, to the formal notion of a ground of a consequence. But God himself would not be constrained in His representation by the concept of causality (because He would not be constrained by our discursive concept of an object in general). Thus, a divine intellect would *not* represent free agency in causal terms. This is a welcome result. That we are constrained to think of freedom in causal terms as a *ground of some consequence* is responsible for our complete inability to comprehend how the causality of freedom works: the notion of an atemporal relation of a ground to a consequence is theoretically elusive (albeit coherent). God is not so constrained: He has a non-inferential understanding of free agency that does not involve the representation of two separable elements (ground and consequence) whose relation would have to be inferentially grounded.

Chapter II: Transcendental Idealism and Transcendental Freedom

Kant's conception of freedom has seemed mysterious to many philosophers. Bernard Williams succinctly describes "the central mystery" as deriving from the fact that Kant

...was a determinist about all events; recognized that actions were events; believed in free will; and condemned the 'wretched subterfuge', as he called it, of making free will and determinism compatible with one another by ascribing free actions to a particular kind of cause.⁷⁴

Williams' impression that Kant's doctrine of freedom is rather mysterious is confirmed by the fact that it has attracted incommensurably different interpretations. Karl Ameriks' complaint that Kant dogmatically rejects compatibilism⁷⁵ sharply contrasts with Hud Hudson's view that Kant anticipated contemporary compatibilist accounts⁷⁶. A third kind of interpretation attributes to Kant a kind of hybrid position; Allen Wood suggests that Kant sought to demonstrate the compatibility of compatibilism and incompatibilism.⁷⁷

In this chapter, I provide an interpretation of Kant's doctrine of transcendental freedom. I argue that while Kant's doctrine is very peculiar, it is also coherent and potentially promising. Its promise lies in its potential to do justice *both* to libertarian intuitions *and* to naturalistic doubts about the theoretical credentials of models of free agency that answer to these intuitions.

1: Interpretive desiderata

I shall begin by describing a number of desiderata that any plausible interpretation of Kant's doctrine must accommodate. First, we must first clarify in what sense Kant is a *determinist*. His determinism concerns everything that happens or exists in time. By that I mean that he accepts the causal principle that every event that has an objective temporal position is causally necessitated by another event⁷⁸, and that the exercise of all causal capacities that we can ascribe to objects on the basis of empirical evidence is governed by deterministic causal laws.⁷⁹ Now,

⁷⁴ Williams 1986a, p. 211.

⁷⁵ Ameriks 2000, p. 207, p. 227.

⁷⁶ Hudson 1994.

⁷⁷ Wood 1984, p. 74. This interpretation is also endorsed by Allison 1990, p. 28 and by Watkins 2004, p. 335.

⁷⁸ With the majority of commentators, but against Watkins 2004, I shall presuppose that Kant's model of empirical causality is that of event-event causation. Watkins claims that the model of causation established by the Second Analogy is one where events are caused by causal powers that are *continuously* exercised. Even if this did have some plausibility as far as the exercise of powers of fundamental physics such as repulsion or attraction is concerned, it would make little sense with regards to cases such as committing a theft which are clearly meant to be within the scope of the Second Analogy (cf. KpV, 5: 95).

⁷⁹ It is, I think, uncontroversial that the exercise of all empirical causal powers is governed by the causal principle and hence subject to natural necessity; cf. CPR, A 204/B 249-50; A 223/B 270. What is controversial is whether Kant believes that the causal principle entitles us to believe that everything that happens is governed by *general laws*. This is the so-called strong, 'same cause, same effect' as opposed to the weak, 'some cause, some effect' reading of the Second Analogy. The strong reading is endorsed by Friedman 1992a and Longuenesse 2000, pp. 143-44; the weak reading is endorsed by Allison 1996, Buchdahl 1974, and Watkins 2004. I think the strong reading is the correct one. In particular, I agree with Friedman that Kant thinks that the law-like character of natural causality follows from the very concept of causality: "for this concept makes the strict demand that something, A, should be such that something else, B, follows from it necessarily and in accordance with an absolutely universal rule" (CPR, A 91/B 123-4). Such an 'absolutely universal' rule governing event-types must be a general causal law. See also CPR, A 536/B 564: "That all events in the sensible world stand in thoroughgoing connection in accordance with unchangeable laws of nature is an established principle of the

Kant thinks that the causal principle has neither the status of an empirical generalization nor that of a conceptual truth. Rather, it is a synthetic a priori proposition whose truth rests on its being a necessary condition of the objectivity of human experience and empirical knowledge (CPR, A 158/B 197; A 197/B 242). Moreover, and very importantly, Kant thinks that one can establish this principle as true *only if* one accepts his doctrine of transcendental idealism. This is, very roughly, the doctrine that space and time are not mind-independent entities but mere forms of human sensibility, and that those forms (in cooperation with our conceptual cognitive faculties) determine conditions both for our knowledge of objects and for the objects of our knowledge.⁸⁰ Kant thinks that the causal principle applies only to *appearances*, namely, to things considered in relation to our forms of sensibility and thus in virtue of their spatiotemporal (empirical) character. It does not apply to *things in themselves* (noumena), that is, to things considered independently of this relation to our cognitive faculties. Unless one accepts this distinction between appearances and things in themselves, one has no grounds for accepting the causal principle as correct: we cannot know universal truths (such as the causal principle) concerning what properties things *must have* independently of their relation to our cognitive faculties, whereas we *can* know what properties things must have in order to qualify as objects of human knowledge. Hence, we can affirm the causal principle as necessarily true if we limit its application to objects of our knowledge, or appearances.⁸¹ Thus, the deterministic causal principle does *not* possess the status of a universal metaphysical truth that pertains to all actions that could possibly be performed or to all causal capacities that could possibly exist.

This must suffice as an explication of the sense in which Kant is a determinist. The second point we need to take into account is that there is very strong textual evidence for the claim that Kant has an incompatibilist concept of transcendental freedom. This 'incompatibilist presumption' can be drawn from the following passage (CPR, A 533/B561-A534/B562):

By freedom...I understand the power of beginning a state spontaneously. Such causality will not, therefore, itself stand under another cause determining it in time, as required by the law of nature. Freedom, in this sense, is a pure transcendental idea. ... That everything which happens has a cause is a universal law, conditioning the very possibility of all experience. Hence the causality of the cause, which itself happens or comes to be, must itself in turn have a cause; and thus the entire field of experience, however far it may extend, is transformed into a sum-total of the merely natural. But since in this way no absolute totality of conditions determining causal relation can be obtained, reason creates for itself the idea of a spontaneity which can begin to act of itself, without requiring to be determined to action by an antecedent cause in accordance with the law of causality.

Transcendental Analytic, and allows of no exception." Allison 1996, p. 86, concedes that the Second Analogy is committed to the idea that there are *rules* governing event-types, but he claims that this leaves open the possibility that there might be only a single token of each event-type and thus that there might be "laws with merely a single instance". This, he claims, would not give us "genuine laws" which require "regularity and repeatability". But suppose there is an object whose constitution is such that whenever conditions c obtain, it must necessarily change its internal state from s1 to s2. Suppose that conditions c obtained once before a change in the atmosphere made it physically impossible for c to obtain again. I do not see why the fact that this law is not repeatable tells against it's being a genuine causal law. We must expect from a (deterministic) law that states, 'if A, then necessarily B' that there could not be an A token without a B token; this does not require or imply the future occurrence of A-tokens. Notice, however, that a 'strong' reading of the Second Analogy leaves open two questions: first, whether the general laws that govern events constitute a *unified system of nature*; and secondly, whether we can *cognize* these laws. This point will become important in section 2 below.

⁸⁰ For a more detailed analysis of Kant's idealism, see chapter I of my dissertation.

⁸¹ This is the point of Kant's 'Copernican Revolution'; see CPR, B XVI-II and A 181/B 223-4.

The concept of freedom delineated in this passages applies to a capacity whose exercise is (a) not causally determined and (b) brings about a new state or event. Conversely, if the exercise of a causal capacity 'stands under another cause determining it in time', this capacity is not exercised freely in the sense delineated in this passage. Since compatibilists characteristically allow that we can apply the concept of freedom to the exercise of a capacity even if this exercise is causally determined, Kant here seems to be delineating an incompatibilist concept of freedom. Compare the following passage (CPR, A 534/B 562):

Obviously, if all causality in the sensible world were mere nature, every event would be determined by another in time, in accordance with necessary laws. Appearances, in determining the will, would have in the actions of the will their natural effects, and would render the actions necessary. The denial of transcendental freedom must, therefore, involve the elimination of all practical freedom.

The implication here is clearly this: if every act were *solely* the effect of empirical capacities whose exercise is governed by laws of nature, there would be no transcendental (and practical) freedom. And this seems to fly in the face of the compatibilist intuition that an action can be called free if it results solely from capacities whose exercise is subject to natural necessity.

The incompatibilist presumption is strengthened by the following passage (KpV, 5: 95-6):

How can a man be called...free at the same moment, and with respect to the same action in which he is subject to an inevitable natural necessity? Some try to evade this by saying that the causes that determine his causality are of such a kind as to agree with a comparative notion of freedom. According to this, that is sometimes called a free effect, the determining natural ground of which lies within the acting thing itself...so although the actions of man are necessarily determined by causes which precede in time, we yet call them free, because these causes are ideas produced by our own faculties... This is a wretched subterfuge with which some persons still let themselves be put off, and so think they have solved, with a petty word-jugglery, that difficult problem, at the solution of which centuries have labored in vain, and which can therefore scarcely be found so completely on the surface.

What Kant refers to here as a 'wretched subterfuge' is the compatibilist strategy of calling an effect – here, an observable act – free if it results from an 'internal' capacity for deliberative choice whose exercise is subject to causal necessity.

Does Kant believe that we *are* transcendently free? At the very least, he affirms what I will call The Rationality Thesis: We can consider ourselves transcendently free.⁸² We will have to see, in the course of this chapter, how precisely this claim is to be understood. Our interpretation of the sense in which Kant confirms the Rationality Thesis is constrained by the fact that Kant also affirms what I will call the No Experience Thesis: Transcendental freedom of the will cannot be an object of experience or theoretical knowledge, and (hence) the causality of freedom cannot serve any role in explaining or predicting actions.⁸³

The importance of these two theses emerges when we take into account the fact that Kant delineates his own position against *two different* opponents. These two opponents can both be regarded as transcendental realists: they both accept that objects of cognition are things in themselves, i.e., that our (veridical) representations exactly correspond to the nature of the object as it is independently of our sensibility. Transcendental realism in this sense is a general 'picture'

⁸² See KpV, 5: 3-4; 28-30 (where it is obvious that the transcendental concept of freedom is at issue) and 47 (where it is claimed that the concept of freedom has 'objective reality', i.e., can be legitimately employed).

⁸³ See, for instance, GMS, 4: 458-9; and CPR, A 542-3/B 570-1.

than can be accepted by various competing doctrines.⁸⁴ The two main opponents of Kant's own view can be described as a dogmatic naturalist (or 'empiricist') and as a dogmatic rationalist.⁸⁵ In the Canon (CPR, A 781/B 809), Kant describes their characteristic kind of error: the rationalist asserts that we have theoretical knowledge that certain concepts, such as freedom, are instantiated even though experience cannot, on Kant's view, yield evidence for their instantiation; whereas the naturalist collapses principles that constrain the constitution of natural objects (appearances) into principles that constrain the constitution of objects in general (things in themselves). The rationalist, as Kant envisages her, rejects the No Experience Thesis, i.e., she applies the concept of transcendental freedom to natural objects and events and makes it serve an explanatory role: "Causality in accordance with laws of nature is not the only causality from which the appearances of the world...can be derived. To explain these appearances it is necessary to assume that there is also another causality, that of freedom" (CPR, A 445/B 473). The naturalist or empiricist, as Kant envisages her, rejects the Rationality Thesis: the naturalist believes that the causality of nature is the only kind of causality that exists and, hence, that we are incapable of causally unconditioned activity: "There is no freedom; everything in the world takes place solely in accordance with the laws of nature" (CPR, A 445/B 473).

Kant agrees with the naturalist against the rationalist that we cannot have experience of transcendental freedom and that natural events cannot be explained by reference to the causality of freedom. Against the naturalist, Kant holds that we *can* consider ourselves free. Furthermore, Kant thinks that we can consider ourselves free *only if* we accept his doctrine of transcendental idealism. A realist, who denies the distinction between appearances and things in themselves, is *not* entitled to the Rationality Thesis (CPR, B XXVI-XXVIII; A 536/B 564; KpV, 5: 95-6).

A final desideratum derives from the fact that Kant accepts that "freedom and nature, in the full sense of these terms, can exist together, without any conflict, in the same actions" (CPR, A 541/B 569) and that "[the deterministic causality of] nature does not contradict the causality of freedom" (CPR, A 599/B 587). These look like claims that a compatibilist would cheerfully

⁸⁴ Here I agree with Allison 2004, p. 28; p. 35; pp. 394-95; p. 445.

⁸⁵ See CPR, A 466/B 494 for the difference between empiricism and the 'dogmatism of pure reason'. For my general understanding of the terms 'naturalism' and 'empiricism', see my introduction, pp. iv-v. I am not claiming here that in Kant the term 'empiricism' denotes one clear-cut doctrine. As far as I can see, Kant uses this term in relation to at least three different doctrines: first, Hume's skeptical view that we do not know whether events really are causally connected; second, Berkeley's view that there are no material substances; and third (in the Dialectic) the (Hobbesian) view that space, time and deterministic causality delimit the bounds of possible existence. My interpretation here is at odds with that of Abela 2002, who holds (p. 28) that Kant's *one* empiricist opponent in the First Critique claims that things in themselves are 'mental primitives' such as impressions and ideas. This characterization may be suitable (although I'm not convinced that it is) with regard to what Kant sometimes (CPR, A 376-7) calls 'empirical idealism', a position which doubts or (cf. Berkeley) denies the existence of material substance. But it should be clear that this characterization fails to capture the spirit of the position that underlies the empiricist antitheses of the Antinomies, which dogmatically declares that everything that exists must be in space and time. Abela suggests that Kant's target in the Dialectic is (p. 220) the assumption, which Abela takes to be shared among the rationalist and the empiricist realist, that general rather than transcendental logic is a sufficient guide toward empirical reality; and he takes (pp. 222-3) Kant to attack a putative empiricist-Humean view according to which whatever is conceivable is experientially possible. But this strikes me as a rather gross misinterpretation. Kant's discussion in the Canon (CPR, A 781/B 809) makes clear that his objection to (one form of) empiricism is that it misinterprets the principles of Kant's transcendental logic as principles of being as such; and the empiricist in the Third Antinomy holds (CPR, A 447/B 475) that freedom is only conceivable but *not* therefore experientially possible (i.e., that it is *merely* "an empty thought-entity").

endorse; they seem to imply a 'compatibilist presumption' that appears to be in sharp tension with what I called the incompatibilist presumption.

I think that the most crucial task for an interpreter of Kant's theory of freedom is to get clear about the status and the ontological import of the Rationality Thesis. This thesis says that we *can* consider ourselves free; if the concept of freedom at issue here is an incompatibilist one that designates the absence of causal determination, the Rationality Thesis appears to license the belief that we are free from causal determination. This seems to conflict with Kant's commitment to determinism, with his No Experience Thesis, and with the compatibilist presumption. Moreover, since Kant thinks that accepting his idealism is required for affirming the Rationality Thesis, the idea that this thesis involves a commitment to our metaphysical freedom from natural necessity must be denied by those who believe that Kant's idealism, in particular his talk about things in themselves, is not to be understood in metaphysical terms.⁸⁶

These considerations motivate the various attempts to 'deflate' Kant's affirmation that we are free agents (i.e., the Rationality Thesis) that one finds in the literature. With respect to those who favor a deflationary interpretation, we need to distinguish between two different strategies: first, the attempt to explain away the incompatibilist presumption and to read Kant as a straightforward compatibilist; second, the attempt to show that the Rationality Thesis does not license a 'theoretical' attitude, i.e., a *belief* in freedom from causal necessity. I will reject the first of those options in section 2 and the second option in section 3 below. This will jointly motivate the claim that Kant's doctrine of freedom involves a commitment to the *belief* that we *are* free from causal necessity. I will then show how this commitment coheres with the compatibilist presumption and develop my overall interpretation of Kant's position on this basis.

2: Against the Davidsonian appropriation of Kant's conception of freedom

A systematic attempt to reconstruct Kant's concept of transcendental freedom in purely compatibilist terms has been proposed by Ralf Meerbote and Hud Hudson, who interpret Kant's theory as an anticipation of Donald Davidson's anomalous monism. Davidson himself represents his theory of action as being "in sympathy with Kant"⁸⁷; Meerbote and Hudson attempt to work out a cohesive interpretation of Kant's account on a Davidsonian basis. Anomalous monism claims: physical and mental events stand in the relation of token-identity; mental events are causally determined since they are token-identical with causally determined physical events; physical and mental events are not type-identical, because mental events can be described by concepts that are irreducible to concepts used by physical theories; there are no deterministic psychological laws, and there are no psycho-physical bridge laws through which we could conceive of mental occurrences as being causally determined under a mental description.

The Davidsonian claim is that Kant's distinction between appearances and things in themselves tracks a difference between two ways or levels of describing one entity or event-token, namely, in physical or in irreducibly mental/psychological terms. When Kant attributes acts to the causality of freedom, he does not mean that the act is produced by a cause that is metaphysically free from causal determination. His talk about noumenal causation refers not to a distinctive way of *producing* action, but to a distinctive, i.e., psychological manner of *describing*

⁸⁶ These philosophers include Allison 1990, 1996, 2004; Friedman 1996; Hudson 1994; Kitcher 1999; Korsgaard 1996; and Matthews 1969. I argue against such views in chapter I, but I do not presuppose these arguments here.

⁸⁷ See Davidson 2001, p. 207.

empirical, physical efficient causation. As Hudson puts it:

Empirical, causal determination [is] the only case Kant allows as a productive or efficacious usage of the term... an agent's actions are determined by his belief in the proposition expressing a relation of means to some end which is an object of desire for that agent. This sense of determination, however, does not appear to be of the nature of a competitor to the sense of 'determination' applied to natural events under the scope of the Second Analogy.⁸⁸

On this reading, Kant holds that every ordinary human action is token-identical with an event that is the effect of a physical cause, but it differs from this event in type: it can be conceived through an irreducibly non-physical description and hence "be regarded as being determined by the causality of reason in [a] sense of pro-attitude, propositional determination". This allows us to speak of "spontaneity as action proceeding from an internal principle (its determination being through the reasons the agent has represented...to herself)".⁸⁹

Now, I do not deny that endorsing this account may seem like a straightforward way of making sense of the fact that Kant thinks that the causality of freedom does not conflict with the deterministic causality of nature even though the concept of freedom has non-deterministic connotations. The trick, on the Davidsonian interpretation, is to regard natural, deterministic causality as the only type of efficient causality but to insist that we can talk about deterministic causes in a manner that abstracts from their deterministic character.

But the Davidsonian interpretation faces insuperable interpretive obstacles. To begin with, the claim that "empirical, causal determination [is] the only case Kant allows as a productive or efficacious usage of the term" is plainly false. The final section of the *Groundwork* abounds with references to the productive causality of freedom: for instance, Kant says that through the idea of freedom we conceive of ourselves as "a priori efficient causes" (GMS, 4: 450). Elsewhere, he refers to transcendental freedom as "the power of beginning a state spontaneously" (CPR, A 533/B 561) and (CPR, A 544/B 572) as "an original act [that] can by itself bring about what did not exist before". Most strikingly, Kant considers transcendently free practical reason "as it is in itself the cause *producing*" actions (CPR, A 550/B 578).⁹⁰

Now, as we saw, Kant (in the 'wretched subterfuge passage' quoted above) rejects the notion that an action can be called free insofar as it derives from a psychological, causally determined cause. This seems to tell against the idea that Kant thinks that describing a causally determined choice in psychological or mental terms is sufficient to consider the effect of this choice as a free action. However, Hudson thinks that this point actually *strengthens* his interpretation, because he thinks that Kant in the 'wretched subterfuge' passage objects to a kind of reductive materialism that mistakenly seeks to reduce mental causes to physical causes. That is: Hudson thinks that Kant's opponent in this passage is someone who treats mental causes as phenomena rather than noumena, and who mistakenly construes mental causation in temporal,

⁸⁸ Hudson 1994, pp. 42-3.

⁸⁹ Hudson 1994, p. 48. Compare Meerbote 1984a, p. 141.

⁹⁰ Proponents of the Davidsonian interpretation believe that if Kant did allow the 'intelligible' causality of reason to be an instance of efficient causality, he would also have to posit law-like correlation between an agent's having a reason to do something and the action (the effect); see Hudson 1994, pp. 49-50 and Meerbote 1984b, pp. 62-3. This worry betrays a misapprehension of the fact that Kant has two concepts of causality: a schematized and an unschematized notion. Both of these are concepts of efficient causation, but only the schematized concept of causality, which applies to empirical causes in the phenomenal world, allows for positing law-like connections.

deterministic terms.⁹¹ This interpretation strictly depends on the following two related ideas: (1) For Kant, talk about noumena involves mental or psychological descriptions, whereas talk about natural phenomena involves physical (materialistic) descriptions; (2) For Kant, natural phenomena cannot be regarded as temporal and as causally determined under psychological descriptions. It is only if (1) and (2) are true that Kant can be interpreted as holding that our freedom rests on the possibility of giving non-deterministic, noumenal, mental descriptions of the physical, causally determined phenomena that compose the natural world of our experience.

(1) must be rejected. For Kant, *nature*, the domain of appearances, comprises both objects of outer and of inner sense; and a doctrine of nature that concerns only objects of inner sense "considers thinking nature"⁹². Hence, Kant accepts that predicates signifying mental states relate to objects of nature, i.e., appearances⁹³: psychological descriptions do not as such refer to things in themselves. Moreover, Kant's notion of a thing in itself applies to objects, such as inanimate bodies, that cannot be described psychologically. So talk about things in themselves is not, as such, tantamount to talk involving psychological descriptions. Hence, the psychological/physical contrast does not map onto the distinction between things in themselves and appearances.

Hudson's evidence for (2) solely consists in an appeal to the skepticism Kant voices (in the *Metaphysical Foundations of Natural Science*) about the possibility of a psychological science.⁹⁴ But we can infer very little from this skepticism. Most importantly, we cannot infer that Kant thinks that things that appear merely through inner sense (i.e., psychological states and events) are not causally determined. Psychological states, such as representations, appear in time, the form of inner sense⁹⁵, and *simply as such* are subject to the 'mechanism' of natural necessity, which does not make them identical to material states⁹⁶:

...the necessity of events in time, according to the law of causality, may be called the mechanism of nature, although we do not mean by this that things which are subject to it must be really material machines. (KpV, 5: 97)

This shows that Kant's skepticism about empirical psychology does not support (2). In conjunction with the rejection of (1), we can say that for Kant the mental is part of nature and *under this description* subject to deterministic causality. So, the crucial point that tells against the Davidsonian interpretation is that for Kant the mental or psychological is not 'anomalous'.⁹⁷

⁹¹ Hudson 1994, pp. 50-55.

⁹² See the Preface to the *Metaphysical Foundations of Natural Science* (Man, 4: 467).

⁹³ Kant argues that "our inner experience...is possible only on the assumption of outer experience" (CPR, B 275). Outer experience is consciousness of physical phenomena, and inner experience is consciousness of mental representations. In the same vein, Kant's rejection of rational psychology leads to the conclusion that "nothing is left for us but to study our soul under the guidance of experience" (CPR, A 382), which implies a commitment to the experience of mental phenomena in accordance with the schematized categories.

⁹⁴ Hudson 1995, p. 67.

⁹⁵ See CPR, A 33/B 49-50: "time determines the relation of representations in our inner state."

⁹⁶ Here I agree with Allison 1990, p. 34. We disagree about what this implies; see footnote 99 below.

⁹⁷ I want to comment on a further aspect of the Davidsonian strategy, which is recommended by Friedman 1996, p. 439, p. 448 and laid out in detail by Meerbote 1982, 1984a, 1984b. The idea is that Kant reconciles freedom and determinism in the *Critique of Judgment* when he shows the compatibility of teleological and mechanistic explanation. This idea lacks textual support, and it suffers from the following crucial difficulty: Kant cannot believe in a tight connection between teleological explanation and freedom, because the teleological model applies even to blades of grass. Even the Davidsonian's favored model, explanation by appeal to *belief-desire* pairs, applies to higher animals. For Kant, free agents are capable of acting in a way whose rationality is *not*

But we can go a step further here. Kant's worries about the potential of empirical psychology relate to our ability to cognize psychological laws and to predict mental phenomena on their basis. One cannot, on this basis, attribute to Kant the view that there are no (deterministic) psychological laws unless one attributes to Kant a verificationism which makes the possibility of scientific prediction of certain phenomena a condition on the existence of deterministic laws governing those phenomena. I do not see any basis for attributing such verificationism to Kant. Rather, I think that Kant is committed to the notion that there *are* psychological laws governing mental phenomena. After all, he mentions (CPR, B 152) "the productive imagination...whose synthesis is entirely subject to empirical laws...of association" which "falls into the domain...of psychology". Even more importantly, Kant claims that

...if we could exhaustively investigate all the appearances of men's wills, there would not be found a single human action which we could not predict with certainty and recognize as proceeding necessarily from its antecedent conditions.... (CPR, A550/B578)

...if it were possible to have so profound an insight into a man's mental character as shown by internal as well as external actions as to know all its motives, even the smallest, and likewise all the external occasions that can influence them, we could calculate a man's conduct for the future with as great certainty as a lunar or solar eclipse.... (KpV, 5: 99)

Kant thinks that our insight into the psychological realm cannot be 'profound' enough to support such prediction of actions. But he also states that if our insight were less limited, we would be able to make the relevant deterministic predictions.⁹⁸ This implies that the relevant laws exist and that our lack of insight prevents us from cognizing them; the verificationist reading would have to claim that our lack of insight into psychological causes implies the absence of laws governing such causes, which seems absurd.⁹⁹

I conclude that the attempt at a compatibilist reconstruction of Kant's concept of transcendental freedom along Davidsonian lines fails because we cannot read Kant as an 'anomalous monist'. I want to add that even if Kant *did* accept some version of anomalous monism, I find it very hard to believe that he would take much comfort in Davidson's evasive

reducible to the dimension of means-end rationality captured by an explanation of activity that appeals to purposes. Hence, the contrast between teleological and mechanistic explanation is a contrast *within* nature, whereas freedom for Kant falls *outside of* nature and cannot explain anything, teleologically or mechanistically.

⁹⁸ Note that such predictions could not be made on the basis of *material* laws. Here I disagree with Amerik's claim (2000, p. 38) that for Kant "the course of our mental life can be correlated with and therefore predicted via reference to fully determined physical phenomena". Kant is an *empirical dualist* who claims that mind and body are different appearances (CPR, A 379-80) and that their changes are governed by irreducibly different laws (CPR, A 683-4/B 711-2). I will explain the interpretive significance of this point in chapter III.

⁹⁹ The verificationist interpretation is rejected by Hatfield 1990, pp. 68-70 and 1992, pp. 217-23. It is (implicitly) encouraged by Allison 1990, pp. 31-4 and Bojanowski 2006, p. 172 ff. Allison's argument is very puzzling: he concedes that what Kant says in the passages I quoted (CPR, A550/B578; KpV, 5: 99) implies that there are strict psychological laws, and he then invokes his weak reading of the Second Analogy, as supporting only the 'some event-some cause' principle (see footnote 79 above). I conjecture that Allison's point here is that the weak reading shows that we cannot take what Kant says in the passages I quoted at face value. *If* this is his point, my response is that the fact that Kant commits himself to the existence of psychological laws (in the passages I quoted) is explained by the stronger reading, but remains wholly obscure on the weaker reading of the Second Analogy; so what Kant says in these passages yields further evidence for the untenability of the weaker reading. According to the stronger reading, it follows from the fact that psychological phenomena are subject to the Second Analogy that they are governed by universal deterministic laws. It does *not* follow that these laws and phenomena form a systematic unity, *nor* that we can cognize these laws either individually or systematically.

idea that we can save our freedom by showing that we can give a non-deterministic *description* of events that as a matter of metaphysical fact are causally determined.

3: The indispensability of the belief in freedom from causal necessity

After rejecting the attempt to deflate the incompatibilist connotations of Kant's concept of freedom, we must now consider the second option for those who claim that Kant's doctrine of freedom involves no metaphysical commitments: a deflation of the Rationality Thesis which says that we can consider ourselves free. I attribute this strategy to Henry Allison and to Christine Korsgaard. (I will refer to the 'A/K account'.) They accept that Kant's notion of freedom cannot be understood along compatibilist lines.¹⁰⁰ But they think that Kant's appeal to freedom is such that (as Korsgaard puts it) "nothing requires any ontological claims"¹⁰¹, or (in Allison's words) that it requires "no ontological conclusions regarding the absolute spontaneity of the self"¹⁰².

What is the point of applying a concept of freedom that signifies independence from causal determination without making an ontological commitment to the claim that we are free in this sense? Allison thinks that the appeal to transcendental freedom is only meant to give us a way of *conceiving* of ourselves. When Kant appeals to causally undetermined agency, this is supposed to be no more than "a conceptual claim...about the model we are constrained to adopt"¹⁰³ from the standpoint of practical deliberation:

...the transcendental idea of freedom...regulates...our conception of ourselves as rational agents. It does so by providing the conceptual basis for a model of deliberative rationality, which includes, as an ineliminable component, the thought of practical spontaneity.

Kant is...claiming merely that it is necessary to appeal to the transcendental idea of freedom in order to conceive of ourselves as rational (practically free) agents, not that we must actually be free in the transcendental sense in order to be free in the practical sense.¹⁰⁴

This seems to give us a very weak version of the Rationality Thesis: the A/K reading seems to construe this thesis as licensing no metaphysical *beliefs* about our freedom but only the adoption of a conceptual model that represents us as free and that is essential for the purposes of choosing and pursuing practical interests (e.g., pure moral interests).¹⁰⁵

I want to first register a philosophical problem that, I think, afflicts the A/K reading. The problem emerges when we ask: why would an agent see the need to 'appeal to' the idea of her freedom from natural causality? Presumably, this need must derive from her sense that the idea that her agency is exclusively governed by natural causality threatens some essential feature of her agency. But this anxiety can be cured only if she has some reason to believe that she might

¹⁰⁰ See Allison 1990, pp. 76-84, and Korsgaard 1996a, p. 187.

¹⁰¹ Korsgaard 1996a, p. 183.

¹⁰² Allison 1996, p. 64.

¹⁰³ Allison 1996, pp. 118-9.

¹⁰⁴ Allison 1990, p. 45, p. 57.

¹⁰⁵ See Korsgaard 1996a, p. 162: "the point [of invoking the idea of freedom] is not that you must *believe* that you are free, but that must choose *as if* you were free"; and Allison 1990, p. 248: "Kant's position is...that freedom is actual, or better, actualized in the interest that we take in the moral law. (...) The claim is not simply that freedom must be presupposed as a condition of the conceivability of such an interest...(although this is certainly part of the story), it is also that our freedom.. is exhibited in our taking such an interest." The elusive idea that our freedom is 'actual in an interest' might allude to a practical rather than a theoretical attitude (i.e., belief) toward the actuality of our freedom. But I am not certain that this is what Allison means to convey here.

actually be free from natural causality. Merely *conceiving of herself as free* from natural causes cannot adequately answer an incompatibilist worry. (Consider an analogy: merely conceiving of the immortality of one's soul cannot cure the anxiety one might feel regarding the idea that 'everything is over' when one's body deceases. An atheist cannot find comfort in forming idle thoughts that carry no degree of conviction or at least hope.) Conversely, if the conviction that one actually is free from natural causality is *not* required for making sense of one's self-conception as an agent, this seems to recommend a compatibilist notion of freedom or agency, rather than Kant's metaphysically ambitious concept of transcendental freedom.

I want to argue that there is an analogous *interpretive* consideration that tells conclusively against the A/K reading. Recall one of the desiderata that I described in section 1: Kant thinks that a naturalist, qua transcendental realist, is not entitled to the Rationality Thesis. The naturalist believes that only the deterministic causality of nature exists. Now, if the idea of transcendental freedom is meant to play a 'merely conceptual role' in our practical deliberation, it seems that a naturalist can happily employ this idea. For she need not discredit the suggestion that we can or should conceive of ourselves in certain terms, as long as this does not involve any *beliefs* that would conflict with her commitment to a naturalistic (and deterministic) ontology.

Now, Allison is of course aware of Kant's insistence that his idealism is required for 'saving' our freedom (i.e., for legitimizing the Rationality Thesis). So how does Allison account for this insistence? At least *one* response is suggested by his discussion of the antithesis of the Third Antinomy. Allison rightly says that "the antithesis represents...a pure empiricism" (what I call 'dogmatic naturalism') and adds that the empiricist "moves from the rejection of transcendental freedom within nature to its complete rejection or, equivalently, from its empirical to its absolute impossibility". By leaving behind empiricism and adopting Kant's idealism one can "find conceptual space for a different conception of causality (transcendental freedom)".¹⁰⁶ Allison's claim that leaving behind empiricism opens up 'conceptual space' for the 'conception of' transcendental freedom suggests that the empiricist assertion that freedom is 'absolutely impossible' means that the *very concept* of freedom is impossible, i.e., contradictory. This suggests that Allison thinks that we must adopt Kant's idealism to make room for the logical possibility of (the concept of) transcendental freedom.¹⁰⁷

But it would be implausible to saddle the naturalist with the position that concepts that do not fit into her ontological framework are logically incoherent. A naturalist need not equate what she deems metaphysically possible with what is thinkable. Moreover, the conclusion of the empiricist antithesis of the Third Antinomy is that "transcendental freedom...is an empty thought-entity" (CPR, A 447/B 475). To say that the idea of transcendental freedom is an *empty* thought-entity is not to hold that it is an incoherent idea, such as the idea of a round square. It is to say that this idea is *merely* thinkable without having any application to reality. Hence, transcendental idealism is not needed to make logical space for the concept of transcendental freedom.¹⁰⁸ The dogmatic naturalist has the exact same conceptual resources as Kant: she can

¹⁰⁶ Allison 2004, p. 383. Compare his 1990, p. 46: "transcendental idealism...is required to create the necessary conceptual space for [the idea of transcendental freedom]."

¹⁰⁷ This is also strongly suggested by his remark (cf. his 1990, p. 52) that with Kant's idealism "...it is (at least logically) possible to attribute...independence [from temporal and causal conditions] to an agent...".

¹⁰⁸ Bojanowski 2006, p. 105 notices that the dogmatic empiricist can think the idea of transcendental freedom without contradiction. But this concession does not prevent him from insisting again and again (p. 14, 16, 50, 53, 79, 85, 122, 170, 184) that the solution of the Third Antinomy shows no more than that we can conceive the idea

talk about noumena and unconditioned causes, and she can invoke these concepts to bolster her self-conception as someone who has practical interests and makes choices. What she cannot allow for is the belief that something actually answers to the conception of ourselves as noumena standing outside the natural (phenomenal) order. But this belief is not required by Kant's account on the weak construal of his Rationality Thesis that is envisaged by the A/K reading.

But maybe the A/K reading is meant to encourage a somewhat stronger version of the Rationality Thesis. Maybe what Allison really wants to say is that Kant's idealism makes conceptual space for the thought that we *may actually be* transcendently free, even if we cannot *know* whether we are.¹⁰⁹ That is: Allison might think that the possibility of freedom enabled by Kant's idealism is *not only conceptual but also metaphysical*. This implies what we may call a moderately strong interpretation of the Rationality Thesis, according to which the Rationality Thesis asserts that we have reason to believe that our freedom from causal necessity is an open metaphysical possibility. This interpretation *does* render the Rationality Thesis unavailable to the naturalist, and it is weaker than an interpretation that takes this thesis to claim that we have reason to believe that we *are* free. However, it is considerably stronger than one might have expected given the official, anti-ontological rhetoric of the A/K reading. For on the current version of this reading, Kant's idealism is intended to make ontological space for the existence of freedom, space that is not available to a naturalist. I do not think that proponents of the A/K reading can be happy with this interpretative line, since it asserts that Kant's 'practical standpoint' does not preempt metaphysical assumptions altogether but rather requires a robustly metaphysical belief in what is ontologically possible; this flies in the face of Allison's claim that

Kant is...claiming merely that it is necessary to appeal to the transcendental idea of freedom in order to conceive of ourselves as rational (practically free) agents, not that we must actually be free in the transcendental sense in order to be free in the practical sense.

If Kant thinks that we can consider ourselves as rational agents only if we treat our transcendental freedom as an open metaphysical possibility, he must think that our status as agents metaphysically depends on whether we *actually are* transcendently free. For if there were no such ontological dependence of rational agency on transcendental freedom, there would be no need for agents to consider transcendental freedom metaphysically possible.

So I think that proponents of the A/K reading face a dilemma. Their official abjuration of ontology leads to a weak version of the Rationality Thesis that requires only the logical possibility of conceiving that we are free without any commitment that this conception might actually be instantiated; and the dogmatic naturalist or empiricist is entitled to the Rationality

of transcendental freedom, or from the thought of an unconditioned causality, without contradiction. He says that a transcendental idealist, unlike a realist, 'possesses the transcendental difference of thing in itself and appearance', where this difference 'enables the thought' that a non-spatial thing in itself underlies spatial appearances (p. 136). But unless this 'thought' involves a commitment to there *being*, or at least potentially being, *something that is not spatial*, nothing prevents the empiricist from entertaining it. Kant's phenomena/noumena distinction can surely be understood and conceived by a committed realist.

¹⁰⁹ Allison 1990, p. 44 (my emphasis): "...by treating space, time, and the categories as epistemic rather than ontological conditions, transcendental idealism also opens up a 'conceptual space' for the nonempirical thought (although not knowledge) of objects...*as they may be* apart from these conditions." However, it is far from clear whether Allison really does mean to suggest here that the appeal to conceptual space is meant to imply ontological possibility. On another occasion (*ibid.*, p. 57) he holds that "Kant is asserting a 'conceptual' rather than an 'ontological' dependence of practical on transcendental freedom".

Thesis thus understood. If they accept a stronger version of the Rationality Thesis according to which the appeal to the conceptual possibility of freedom is at the same time an appeal to metaphysical possibility, then the naturalist is not entitled to the Rationality Thesis; but this shows that Kant's theory of freedom and agency requires strong metaphysical foundations.¹¹⁰

My discussion of the A/K reading has an important upshot: the logic of Kant's position strictly implies that he thinks that his account of freedom *has ontological import*, and that we can consider ourselves free (i.e., that we can affirm the Rationality Thesis) only by virtue of endorsing his idealism (specifically: by rejecting the naturalistic worldview that is implied by the antitheses of the Antinomies). That is: we must read Kant as affirming either that our freedom from natural necessity is an open metaphysical possibility (this is what I called the 'moderately strong' version of the Rationality Thesis) or even that we are in fact free from natural necessity. I now want to confirm, on direct textual grounds, that Kant's account of freedom has ontological import, and on this basis I shall settle the question of how substantive this import is meant to be (i.e., whether Kant accepts the weaker point that our freedom from natural necessity is an open metaphysical possibility or the stronger point that we are free from natural necessity).

We must acknowledge, first of all, that Kant often claims that adopting his idealism has the benefit that "the representation of [freedom] is at least not self-contradictory" (CPR, B XXVIII; see also A 558/B 586; and GMS, 4: 456). This might encourage the impression that Kant accepts the (wholly implausible) idea that his realist opponent is not even entitled to the concept of freedom. But a closer inspection of the textual environment of the relevant passages shows that this impression is false. The above quote is preceded by the suggestion that transcendental realism would imply that "I could not...without palpable contradiction, say of one and the same being, for instance the human soul, that its will is free and yet is subject to natural necessity" (CPR, B XXVII). The propositional attitude ('saying of') toward 'the will is free' envisaged here is the same as that toward 'the will is causally determined'. Since Kant clearly thinks that we should *assert* the proposition, 'the will is (qua phenomenon) causally determined', the attitude toward the proposition 'the will (qua noumenon) is free' that his idealism is supposed to enable must be that of assertion as well. Kant's point here is this: his naturalist opponent believes (a) that all causality is mere nature, and so she would be contradicting herself if she asserted (b) that we also have a capacity for activity whose exercise is not governed by natural laws. If Kant's idealism is accepted, and if the commitment to determinism as a universal metaphysical truth that is inherent in (a) is given up, this allows us to endorse both the claim (c1) that our will is causally determined and the claim (c2) that our will free. When Kant says that his idealism allows that the representation of freedom is not self-contradictory, he does not mean that his idealism enables the mere idle *thought* of freedom (alongside that of The Golden Mountain), but that it removes the threat of a contradiction between (c1) and (c2) and thereby

¹¹⁰ It should be noted that in his 2004, p. 48, Allison concedes that Kant holds that we can *affirm that* we are free (only) from a practical standpoint. I fail to see how this coheres with Allison's 'official' interpretation of Kant's theory of freedom in his 1990, 1996, where (as we saw) Allison tries very hard to do away with the idea that Kant is committed to affirming that we really are free. Maybe Allison thinks that Kant wants to affirm only that we are *practically* free; he thinks (1990, p. 57) that the notion of practical freedom is less metaphysically loaded than that of transcendental freedom, a point that I find completely unpersuasive in view of Kant's claim that "the denial of transcendental freedom must...involve the elimination of all practical freedom" (see p. 31 above; I discuss this point in more detail in chapter III). Or, Allison might think that affirming our freedom from the practical standpoint amounts only to adopting a practical attitude that falls short of full-fledged belief (see footnote 105 above), and/or that carries no ontological commitments.

makes logical space for the belief that we are free.¹¹¹ Likewise, accepting Kant's idealism instead of an exclusively naturalistic-deterministic ontology opens up *ontological space* for the existence of transcendental freedom.¹¹² This is the point of Kant's appeal to his idealism in his solution to the Third Antinomy in the Transcendental Dialectic of the Critique of Pure Reason.

Now, the 'moderately strong' interpretation of the Rationality Thesis holds that we should read Kant as asserting *no more* than that our freedom is metaphysically possible. As we just saw, Kant takes the metaphysical possibility of freedom to be established *already* by the argument in the Dialectic of the Critique of Pure Reason. Hence, the moderately strong interpretation is committed to saying that the Rationality Thesis is established already as a result of Kant's argument in the Dialectic. But this is clearly false. Kant's attitude toward our transcendental freedom fully emerges in the Critique of Practical Reason, and Kant's argument here is intended to build on and add to the argument of the Dialectic.¹¹³ Hence, Kant intends the Rationality Thesis, as it emerges in the Second Critique, to be stronger than what the argument of First Critique permits. This implies the following overall interpretation: the argument of the Dialectic creates ontological space for the existence of freedom and thereby logical space for the belief in freedom; the appeal to the moral law (the 'fact of reason') in the Critique of Practical Reason builds on this foundation by showing that we have positive reason for adopting this belief. Hence, we should interpret Kant's Rationality Thesis as claiming that we have reason to believe that we are free, i.e., that the ontological possibility of freedom is realized in us.

This is only a sketch of the overall logic of Kant's position: I will reconstruct his argument for our belief in freedom (based on our awareness of the moral law) in a later chapter.

¹¹¹ In the Groundwork (GMS, 4: 498), Kant suggests that his idealism removes the worry that "the thought of freedom contradicts...itself", which again seems to support the very weak interpretation envisaged by Allison. But, again, Kant clarifies: "It would...be impossible to escape this contradiction if the subject, which seems to itself free, conceived itself in the same sense or in the very same relation when it calls itself free as when in respect of the same action it assumes itself to be subject to the law of nature." The threat of contradiction is here raised with respect to the conjunction of 'calling oneself free' and 'assuming oneself to be subject to natural laws'; and this implies that the first pair of the conjunction has the same propositional status as the second pair, namely, one of full-fledged assertion. 'The thought of freedom' is hence 'the assertion that I am free'.

¹¹² In his solution to the Third Antinomy, Kant raises the question of "whether freedom is possible at all, and if it be possible, *whether it can exist* along with the universality of the natural law of causality" (CPR, A 536/B 564; my emphasis), or whether freedom "can take place" despite the truth of the Second Analogy; these questions, he suggests, cannot be answered affirmatively by the naturalist qua transcendental realist.

¹¹³ In the Critique of Practical Reason (KpV, 5:47), Kant appeals to the moral law "as a principle of the deduction of freedom as a causality of pure reason...since theoretical reason was compelled to assume at least the possibility of freedom". There is a complication: in the Critique of Pure Reason, Kant says (CPR, A 558/B 586) that theoretical reason does not even show the possibility of freedom. The solution to this puzzle is that he is equivocating on various senses of possibility. Theoretical reason, via the indirect proof for transcendental idealism provided by the Antinomies, does *not* show that the concept of freedom has what Kant calls 'real possibility' or objective reality (cf. CPR, A 218-26/B 265-73). The real possibility of a causal power requires that we have empirical or a priori reasons to believe that this power exists. Kant's idealism *by itself* does not show that we have such reasons to believe in the causality of freedom; it only shows, by denying that determinism is a universal metaphysical truth, that freedom is not logically or (what amounts to the same for Kant's critical conception of modality) metaphysically impossible. The Critique of Practical Reason tries to show that the moral law makes the concept of freedom really possible, or objectively real, in some practical sense, where this means (roughly) that it supplies some kind of practical justification for the belief in freedom. Thus, Watkins' claim (2004, p. 339) that "Kant's entire purpose is to show simply that the impossibility of freedom cannot be demonstrated" is much too weak as an account of Kant's considered stance.

4: Kant's theory of freedom as the compatibility of hard determinism and libertarianism

The result of sections 2 and 3 is that we must understand Kant as holding that we are metaphysically free from determination by natural causes. This suggests that Kant is a 'soft incompatibilist' or libertarian. Yet, as we saw in section 1, Kant is also a determinist. Furthermore, he holds that "[the deterministic causality of] nature does not contradict the causality of freedom" (CPR, A 599/B 587), which seems like a compatibilist view.

To unravel this mystery, I want to note first that it is highly misleading to call Kant a compatibilist. The sense that Kant must be a compatibilist derives from the fact that he accepts that an act's or agent's being causally determined does not tell against the act's or agent's being free. But notice that Kant accepts this *because, and only because*, he accepts that the fact that an act or agent is *unfree* does not tell against the fact that the act or agent is free. For Kant, the relation of compatibility obtains between freedom and causal determination *only* in virtue of the fact that this relation obtains between freedom and its negation. 'Causally determined' is, for Kant, synonymous with 'unfree': he holds that one and the same will can be considered as free and "as subject to natural necessity, that is, as not free" (CPR, B XXVII), and he deems 'causally determined' and 'free' "contradictory concepts" ("einander widerwärtige Begriffe"; KpV, 5: 95). So the term 'compatibilism', as it applies to Kant, differs greatly from the term as it is commonly understood: *the* characteristic compatibilist intuition is precisely that the concepts 'free' and 'causally determined' do *not* contradict one another. Hence, Wood's idea that Kant seeks to show "the compatibility of compatibilism and incompatibilism" is misleading, because it uses a standard notion of 'incompatibilism' but a strictly aberrant notion of 'compatibilism'.¹¹⁴

This raises the question of how Kant can coherently think that the belief that an agent or act is free is compatible with the belief that the same agent or act is causally determined and thus unfree. Here his position is analogous to his view that objects are both spatial and non-spatial:

If we add the limitation of a judgment to the concept of the subject, the judgment is unconditionally valid. The proposition 'All things are next to one another in space' is valid only under the limitation that these things be taken as objects of our sensible intuition. If here I add the condition to the concept and say: 'All things, as outer intuitions, are next to one another in space', then this rule is valid universally... Our expositions...teach the reality...of space in regard to everything that can come before us externally as an object, but at the same time the ideality of space in regard to things when they are considered... without taking into account the constitution of our sensibility. (CPR, A27-8/B 43-4)

If, then, we would attribute freedom to a being whose existence is determined in time, we cannot except him from the law of necessity as to...his actions... Now as this law inevitably applies to all the causality of things, so far as their existence is determinable in time, it follows that if this were the mode in which we had also to conceive the existence of these things in themselves, freedom must be rejected as a vain and impossible conception. Consequently, if we would still save it, no other way remains but to consider that the existence of a thing, so far as it is determinable in time, and therefore its causality, according to the law of natural necessity, belong to appearance, and to attribute freedom to the same being as a thing in itself. This is certainly inevitable, if we would retain both these contradictory concepts together. (KpV, 5: 95).

...there is no contradiction in supposing that one and the same will is, in the appearance, that is, in its visible acts, necessarily subject to the law of nature, and so far not free, while yet, as belonging to a thing in itself, it is not subject to that law, and is therefore free. (CPR, B XXVII-III)

Space and time are forms of human sensibility that determine a generic (phenomenal) character

¹¹⁴ I take my disagreement with Wood here to be verbal rather than substantive; that said, my impression from many conversations with various people is that the idea that Kant is a compatibilist causes a lot of confusion.

for objects of human knowledge ('appearances'), i.e., for anything that we can represent as part of a unified spatiotemporal framework. Kant argues that our cognitive faculties impose conditions both on the experience of object and on the objects of experience (CPR, A 158/B 197). One such condition is the causal principle that all knowable objects are subject to law-like causal necessity (because, Kant argues, representing things as governed by law-like necessity is a condition of assigning them an objective temporal position in one course of experience). Since human agents *are* objects of knowledge, they can be represented in space and time and as parts of nature, and thus we must view them as appearances, as causally determined, and hence as unfree.

However, Kant argues, space and time are *only* forms of human sensibility rather than mind-independent entities comprising everything that exists; and hence (as I pointed out earlier), the deterministic causal principle, as a condition of temporal unity, does not have the status of a universal metaphysical truth. The conditions imposed on objects of our knowledge, including human agents, by our cognitive faculties do not pertain to the constitution that these objects have independently of their relation to our cognitive faculties, namely, in themselves (qua noumena). Since there is no contradiction in supposing that one and the same object both has a property that it owes to a certain relation and lacks that property independently of this relation, there is no contradiction in supposing that we are both (qua appearances, in relation to human cognizers) causally determined and (qua things in themselves) free. Correspondingly, if we consider observable actions as natural events, we must attribute them to causal capacities whose exercise is governed by deterministic laws. But we can also attribute them to causes that are not subject to law-like necessity; considered in this manner, actions are free. There is no contradiction between regarding actions as both free and not free because there is no contradiction between attributing them to both natural and non-natural causes. We can attribute actions to two distinct types of causes because we can consider the agent in two ontologically distinct ways: as an appearance (i.e., as an object whose character conforms to the character of our cognitive faculties), or as a thing in itself (i.e., as an object whose character is independent from the conditions imposed on things by our cognitive faculties, including the condition that objects and their acts be causally necessitated). (See chapter I for a more detailed discussion of these issues.)

Now, the compatibility of freedom and determinism enabled by this picture may seem too abstract to temper the worry that the idea that we are both free and unfree is somewhat schizophrenic. Here Kant would respond that his reconciliation of freedom and determinism by ascribing freedom to our non-empirical character and by ascribing causal necessitation to our empirical constitution *has* a repercussion in ordinary thought: we do not consider ourselves as free and as unfree simultaneously or within the same context. When we have in view the fact that we are causally determined, we do not have in view the fact that we are free, and vice versa. This is what motivates Kant's doctrine of two standpoints, which holds that

...[each rational being] has two standpoints from which he can regard himself, and recognize laws of the exercise of his faculties, and consequently of all his actions: first, so far as he belongs to the world of sense, he finds himself subject to laws of nature...; secondly, as belonging to the intelligible world, under laws which being independent of nature have their foundation not in experience but in reason alone. (GMS, 4: 452)

Kant's point here is that the commitment to our freedom comes to light only in certain contexts, when we face questions that concern the *normative* rather than the *natural* order of things (what *ought to* happen rather than what *does or will* happen) and that do not call for the representation

of causal necessity. Likewise, our commitment to the notion that we are causally determined parts of nature arises only when we take an empirical perspective and address a corresponding set of (explanatory, observational, predictive) questions that do not concern us as free agents.

Now, the appeal to the fact that for Kant the concern with freedom arises only from a non-empirical standpoint is also characteristic of the A/K reading. But for Allison and Korsgaard, the point of Kant's appeal to the practical standpoint is that adopting this standpoint altogether absolves the deliberator from the need for an 'ontological' belief in freedom. This interpretation runs afoul of the worry I raised in section 3. If adopting the non-empirical standpoint requires only that we 'conceive of' ourselves as free and take a practical interest in this conception, we do not need Kant's idealism to save the legitimacy of this standpoint. A naturalist may take a practical interest in a conceptual model she considers inapplicable to reality.¹¹⁵ On my interpretation, Kant's appeal to distinct standpoints *presupposes* that the metaphysical resources provided by his idealism save the compatibility of the two *assertions* (that we are free and that we are unfree) that are associated with these standpoints and hence prevent the two standpoints from contradicting each other. But the appeal to the notion that these two beliefs arise from different standpoints adds something to the appeal to compatibility: it shows that the relevance of the two beliefs is limited to a specific form of enquiry. Given the irreducibly different questions associated with the two standpoints, we are not forced to view ourselves both as free and as unfree in one and the same context. In a context where our focus is on our freedom, our lack of freedom is a non-issue, and vice versa.¹¹⁶ This is only a preliminary account of Kant's standpoint distinction; I discuss the intricacies of this distinction at length in chapter VIII. For our present purposes, the crucial point is that an interpretation that emphasizes Kant's standpoint distinction and an interpretation that places weight on the metaphysical commitments of Kant's theory of freedom are not mutually exclusive.¹¹⁷

I suggest that we should represent Kant's overall position as a synthesis of *libertarianism and hard determinism*. The libertarian stance is the one we must take if we consider human beings as they are in themselves, according to their intelligible (noumenal) character; the hard determinist stance is the one we must take if we consider human beings as appearances, according to their sensible character, as being on a par with all other, unfree natural objects:

¹¹⁵ Matthews 1969, p. 214 claims that Kant's distinction between a 'standpoint of agency' and a 'standpoint of knowledge' preempts all metaphysical concerns about determinism, and he takes pleasure in the 'irony' that Kant's conception of freedom on his interpretation turns out to be strikingly similar to that of P. Strawson (2003), someone who – according to Matthews – mistakenly reads Kant's idealism as a metaphysical doctrine. But the jest is on Matthews, for in assimilating Kant's theory to Strawson's naturalistic compatibilism, he has led Kant's claim that a transcendental realist is not entitled to his doctrine of freedom entirely ad absurdum.

¹¹⁶ One can challenge this idea: for instance, one can hold (cf. Nelkin 2000; Watkins 2004, p. 322) that from the standpoint of freedom the representation of empirical possibility *does* play a role, namely, insofar our awareness of physical possibility constrains our sense of what bodily acts we are capable of performing. I think that Kant would respond that when we adopt the non-empirical standpoint and (e.g.) form practical maxims, we entirely bracket the question of what *inner acts* (of volition or choice) we, considered as phenomenal beings, empirically can perform (see, for instance, CPR, A 548/B 576). After all, Kant's concern is only with freedom of will, not with freedom of body; he holds that it is *only* volitional acts that are always within our control and for which we are primarily responsible (see KpV, 5:36-7 and my chapter VII). I return to this point in chapter VIII.

¹¹⁷ For the view that Kant's standpoint distinction is meant to be an alternative to an ontological reading of Kant's idealism (and, consequently, of his views on freedom), see Abela 2002, pp. 38-9, Allison 2004, pp. 47-8 and Korsgaard 1996a, p. x (in her 'Introduction'). For the view that a metaphysical interpretation of Kant's theory is in conflict with an interpretation that places weight on Kant's standpoint distinction, see Watkins 2004, p. 321 ff.

In its empirical character...this subject, as appearance, would have to conform to all the laws of causal determination. To this extent it could be nothing more than a part of the world of sense, and its effects, like all other appearances, must be the inevitable outcome of nature. (...) In its intelligible character...this same subject must be considered free from all influence of sensibility and from all determination through appearances. (...) And...since natural necessity is to be met with only in the sensible world, this active being must in its actions be independent of, and free from all such necessity. (CPR, A 540-1/B 568-9)

Kant's concept of transcendental freedom is, as we saw, robustly incompatibilist in that it requires the absence of causal determination. Kant deems this concept *inapplicable* to objects of human sensibility, i.e., to objects as viewed from the empirical standpoint, considered according to their phenomenal constitution: when we consider objects and their acts as parts of the mechanistic, spatiotemporal order of nature, reason *prohibits us* from representing them as free or spontaneous (CPR, A 542-4/B 570-2) and requires us to say of any act that actually occurred in time that it was impossible that the act could have been omitted (KpV, 5: 95), which removes the basis for judgments of responsibility and for prescriptive ought judgments: "When we have the course of nature alone in view, '*ought*' has no meaning whatsoever" (CPR, A 547/B 575). This is hard determinism in a most uncompromising form. Our entitlement to the belief in freedom presupposes the metaphysical legitimacy of a perspective from which we do *not* consider human agents as part of the mechanistic order of nature and from which we bracket the standards that are constitutive of our attempts to represent and to understand the natural world: namely, a perspective from which we consider human beings according to their noumenal character. From this perspective, we abstract from the properties (such as being causally necessitated) that objects possess in relation to human sensibility – thus, from this standpoint, we *take determinism to be false* as a claim about noumena (we consider deterministic causality to be 'transcendentally ideal'). This is a form of libertarianism.

Kant thinks that the libertarian and the hard determinist perspectives are compatible (i.e., they do not contradict one another). But, to repeat, it would be misleading to infer that therefore Kant can be considered a 'compatibilist' in the common sense of that term. A compatibilist holds that there is *one* perspective on human action from which we can recognize *both* that the action is causally determined *and* that the action is free (e.g., that the agent would have acted differently if she had chosen to do so). For Kant, the claim that an act is causally determined implies that the act is unfree; attributing our actions to free causes requires a standpoint from which we can deny that we are causally determined. Freedom and determinism are never part of one and the same story about human action.

5: Problems and prospects for Kant's conception of freedom

On my interpretation, Kant thinks that we can combine the belief that an act is causally determined with the belief that the act results from a cause whose causality is not governed by natural laws and that operates outside of time and nature. How can we make sense of the notion that an action is the effect *both* of a natural, deterministic cause *and* of a cause whose causality is not constrained by deterministic laws? And how can we make sense of the notion of a cause that operates independently of temporal conditions?

Here it is important to get clear about the demand that we must be able to *make sense of* these notions. Thomas Nagel dismisses Kant's position on the grounds that it is

"unintelligible".¹¹⁸ But this can mean two things. First, it can mean that Kant's position is *incoherent*. And this is false. The idea of timeless, uncaused agency is not like the idea of a round square, because (as the concept of divine agency shows) it is not part of the concept of agency that agency must occur in time. Kant's appeal to timeless agency is predicated on the assumption that time is *only* a form of human sensibility rather than an absolute metaphysical constraint on objects and activities. Likewise, the idea of an action resulting both from determined causes and from non-determined causes is coherent, because one can consistently believe of an action that it stands both under a deterministic and a non-deterministic causality – if one gives up the notion that determinism is a metaphysical constraint on all possible activity and instead views it as a concept that pertains only to the empirical constitution of agents, namely, to the causal powers they possess as spatiotemporal objects of human knowledge.¹¹⁹

But the claim that Kant's conception of freedom is 'unintelligible' can also mean that this conception is *theoretically unsatisfactory, even profoundly so*, because it invokes a causal model that strains our imagination and that has no scientific credibility. And this is undeniably correct. It is hard to get our mind around the model Kant envisages, and the model is at odds with the notion of causation we employ to explain things that happen. Commentators often try come up with some story that makes Kant's model more tractable: suggestions range from the idea that our "timeless choice of an intelligible character spontaneously [determines] each individual act as that act occurs in time"¹²⁰ to the idea that what we timelessly choose are the laws of nature which govern our phenomenal choices and actions.¹²¹ But these ideas are as elusive as the model that they purport to make sense of. I suggest that we remain more faithful to Kant's intentions if we do not attempt to fill in the extremely vague contours of his model of free action. Any such attempt would try to give Kant's model an air of determinacy that it cannot possibly have in Kant's system *by his very own admission*. Kant explicitly affirms that we cannot understand how the causality of transcendental freedom operates: so the idea of transcendental freedom *is* indeed to be regarded as theoretically unintelligible. This is part of the point of what I referred to as the 'No Experience Thesis'. Kant's account of freedom is not aimed at giving us a *theory* of action that allows for an explanation of what happens when someone acts, or at giving us a notion of substance or agent causation that might rival the explanatory resources provided by the naturalistic model of event-event causation.¹²² For Kant, speculation about how freedom operates

¹¹⁸ Nagel 1989, p. 119. Compare Guyer 1987, pp. 303-5, and Strawson 1966, especially pp. 235-72.

¹¹⁹ This point is important to assess one of Bennett's objections to Kant's account (1974, pp. 199-200; for a similar worry, see Beck 1987, p. 42). He points out that Kant holds (a) that our actions are *not only* subject to natural causality but does not dispute (b) that *all* our actions are subject natural causality; and he argues that only rejecting (b) would provide comfort for an incompatibilist about responsibility. But the reason why Kant thinks that an incompatibilist can find comfort in (a) is that he would not accept the belief that Bennett attributes to him, "natural determinism is true", without qualification. "The icy grip" of natural causality only reaches our phenomenal constitution; in allowing that all our phenomenal acts are subject to natural causality, Kant is relying on his idealist understanding of natural causality (i.e., of the claim that we are causally determined) according to which natural causality is 'transcendentally ideal', namely, literally *nothing* independently of our cognitive faculties. This makes room for the idea that our phenomenal acts are metaphysically grounded in noumenal acts that are not subject to natural causality. In other words, Kant finds comfort in (a) because he does not accept that an appeal to natural causality says something metaphysically fundamental about the grounds of our phenomenal (observable) actions (although we can appeal only to natural causality if we occupy the empirical standpoint).

¹²⁰ Wood 1984, p. 96.

¹²¹ Watkins 2004, p. 334.

¹²² Watkins 2004 (see especially p. 411) tries to reconstruct Kant's conception of freedom as an explanatorily potent

and how it interacts with deterministic natural causes in producing observable events must be wholly futile, fruitless and pointless.¹²³ He would also insist that the attempt to *understand how the causality of freedom operates* betrays a misapprehension of the fact that our interest in our freedom arises from a standpoint that brackets theoretical, explanatory questions.

Here we must be careful: we must not conflate the (correct and important) point that for Kant the idea of freedom does not lend itself to metaphysical theorizing aimed at specifying how free causes operate with the (incorrect) suggestion that invoking this idea requires no ontological assumptions about what causal capacities we in fact possess. The necessary lack of success of attempts to make Kant's model of free agency theoretically tractable should not encourage the idea that Kant abjures a metaphysical belief in freedom, but rather the idea that Kant admonishes us not to treat this belief like a belief in ordinary empirical capacities that we can understand on the basis of perception (GMS, 4: 458). The assumption that we are transcendently free involves a commitment to the claim that we have non-natural causal capacities. Theoretical reason, in articulating the distinction between appearances and things in themselves, devises a metaphysical picture within which the apparently conflicting commitments in our self-conception (to freedom and determinism) can be reconciled. Beyond this, theoretical reason has no metaphysical insight.¹²⁴ The ontological commitment to our freedom that reason makes room for must remain theoretically indeterminate since the idea of transcendental freedom lies beyond the bounds of objective sense experience (albeit not beyond the 'bounds of sense'). In other words, while it is important for Kant to make metaphysical space for the idea that we have the capacity for absolutely spontaneous action, it is equally important for Kant's critical position that this space cannot be determinately filled on the basis of intuitive insight or conceptual analysis.

As I argued in chapter I, Kant takes his idealism to show quite generally that objects have an ontological character that differs from the the character that they owe to their relation to our

model of substance or agent causation (remember, however, that he does not accept that Kant holds a naturalistic model of event-event causation even at the phenomenal level; see footnote 78 above). The theory he comes up with is ingenuous in many ways, but I do not see that anything in Kant answers to or calls for it. Watkins presents his agent causation model as a response to the worry, raised by C.D. Broad and others, that the idea of agent causation is 'incoherent'. But, as we saw, the claim of incoherence is ambiguous: it can mean either (a) that the idea of agent causation is self-contradictory or (b) that this idea is theoretically fruitless. At one point, Watkins suggests (p. 412) that his argument shows that "the very idea of agent causation is thus not incoherent", which suggests that he is responding to (a). But (a) is *clearly* false: there is nothing in the idea of agent causation that threatens to put it on a par with the non-concept of a round square. Now, at another point (p. 410), Watkins suggests that the worry he is responding to is that "proponents of agent causation can give no explanation of why the effect that the agent brings about" occurred at a given time rather than at any other time. This suggests (b), and I think that Kant would wholeheartedly concede this: i.e., he would admit that his conception of free agency cannot *explain why* acts occur at the time they do or, indeed, why they occur at all. Even in the Second Critique, where Kant is (arguably) most optimistic that he can demonstrate that we *are* free, he emphasizes over and over again that the concept of freedom is of no theoretical and hence of no explanatory use, and that we cannot theoretically comprehend free agency (KpV, 5: 49-50; 54-55; 98-99). We should also keep in mind that Kant associates the idea that the idea of freedom can explain observable actions with the rationalist position of the Third Antinomy (CPR, A 445/B 473), which he rejects.

¹²³ On this point, I am in full agreement with Korsgaard 1996a, p. 203.

¹²⁴ This line of thought is explicit in the Groundwork: "Hence it is an indispensable problem of speculative philosophy to show that its illusion concerning the contradiction rests on this, that we think of man in a different sense and relation when we call him free and when we regard him as subject to the laws of nature as being part and parcel of nature." (GMS, 4: 456). "But reason would overstep all its bounds if it undertook to explain...how freedom is possible" (GMS, 4: 458-9).

cognitive faculties, and that is *therefore* theoretically incomprehensible to us. His conception of freedom exemplifies this general point. When James van Cleve says, about atemporal causation, "I do not say that such a relation is unintelligible or incoherent, but I do find it difficult to comprehend"¹²⁵, his statement is (unwittingly) congenial to Kant's intents and purposes.

So Kant himself would not be bothered by the worry that his appeal to the causality of transcendental freedom implies a theoretically fruitless model of causation. But a sympathetic reader of Kant cannot simply rest content with embracing this as an unavoidable upshot of Kant's system. This is so for the following reason. Kant thinks that the fact that his idealistic system makes room for the belief that we are free is at least *part of* what recommends this system over realist alternatives.¹²⁶ So one cannot simply take this idealistic system *as given* in assessing whether the price of theoretical incomprehensibility that it imposes on the conception of freedom that it enables is worth paying. To motivate the idea that it is *worth paying* the price of theoretical fruitlessness imposed by Kant's ambitious concept of freedom, one must show that the only alternative to accepting Kant's metaphysically ambitious conception of freedom is a shallow *wretched subterfuge*. If this could be shown, Kant's position might be attractive as a 'last resort'.

I think that Kant generally considers his idealism such a last resort: he thinks that it is the only way to save, among other things, our knowledge of mathematics (CPR, A 48-9/B 66), our entitlement to the notion that natural events are connected through law-like necessity¹²⁷, and the ideas of freedom, morality, and responsibility. With respect to any of these issues, one's willingness to accept Kant's costly idealist solution hinges on whether one is convinced that more conventional and less costly ways of treating the issue at hand are hopeless. Thus, with regards to freedom, morality, and responsibility, Kant's rejection of compatibilism plays an essential role. If Kant's rejection of compatibilism were, as some commentators claim, based on nothing but "a mere dogmatic attachment"¹²⁸, the compatibilist could point out that her conception of freedom is both practically adequate and, unlike Kant's conception, theoretically satisfying. Now, Kant himself indeed did not spell out why, precisely, his metaphysically ambitious conception of freedom qua absolute spontaneity is essential to our self-conception as agents in a way that cannot be simulated by compatibilist alternatives.¹²⁹ However, this leaves

¹²⁵ Van Cleve 1999, p. 138.

¹²⁶ Kant considers the fact that his idealism resolves the Antinomies of reason, including the conflict between freedom and determinism, an *indirect* proof of his idealism (CPR, A 507-8/B 535-6; compare the footnote at B XIX). He feels entitled to presuppose his idealism in his solution to said conflict because he thinks that he has already given a *direct* proof for his idealism in the Transcendental Aesthetic, where he argued that assuming the ideality of space and time is the only way to account for the a priority of mathematical knowledge.

¹²⁷ As a reminder: this is because, Kant argues, the principle that events are connected through natural necessity can only be established as true if (a) it can be shown that it is a condition of the possibility of experience and (b) if one accepts that objects of experience are not independent of the conditions of human experience. See CPR, A 181/B 223-4. This is also the point of Kant's remark that if transcendental realism is accepted, *neither nature* (as a system of events connected through necessitating laws) *nor freedom* remains (CPR, A 543/B 571).

¹²⁸ Ameriks 2000, p. 227.

¹²⁹ This can be seen by considering Kant's 'official' rejection of compatibilism (KpV, 5: 94-7). The compatibilism he attacks here intends to track a notion of freedom of action that signals the absence of coercion through external forces; for example, if A hits B in the face, this action can be regarded as free if the action is caused by the agent's own choice, but not if there is (say) a tornado that tosses A against B. Kant's derisive rejection of such compatibilism lends itself to the following argumentative mode: (1) if we are free in the compatibilist sense, so is a turnspit; (2) a turnspit is not free in any relevant sense; (3) so we are not free in any relevant sense. The crucial assumption in this *modus tollens* is obviously the first premise. Compatibilists would reject it on the grounds that

open the possibility that one can reconstruct, from his often cryptic remarks, forceful incompatibilist arguments that motivate his theoretically fruitless conception of freedom.

There is a second reason why simply calling attention to the fact that the theoretical fruitlessness of Kant's concept of freedom is an unavoidable upshot of his idealist system cannot by itself defuse worries about the credentials of this concept. It is very important to see that Kant's idealism by itself does not give us any *positive* reason for the belief that we *are* free. Transcendental idealism shows that our noumenal character differs from our empirical character and that the principle that actions follow necessarily from preceding events according to empirical laws places a constraint *only* on the exercise of our empirical powers. This makes ontological space for freedom, and it allows one to believe without contradiction of one and the same object that it is (in virtue of its non-empirical character) free and (in virtue of its empirical character) unfree. But the fact that we can believe that we are free without contradiction does not show that we have any reason to believe this. That is: from the fact that Kant's idealism establishes the notion that we have an ontological constitution that differs from our empirical constitution we cannot infer, without further argument, *that* our non-empirical constitution comprises the capacity for transcendental freedom.

One might object that it just follows from (1), 'The principle that all activity is causally necessitated constrains only our empirical capacities' and (2) 'We *have* an ontological constitution that differs from our empirical constitution' that (3) 'Our non-empirical character comprises causal powers whose exercise is not constrained by empirical laws and conditions'. But the inference to (3) is fallacious, and clearly rejected by Kant. Kant does not, for instance, infer from the fact that we, as appearances, are mortal that we, as things in themselves, are immortal (CPR, A 394-5). Likewise, it does not follow from the fact that our noumenal character is not to be described in terms of empirically conditioned causal capacities that this character comprises empirically unconditioned capacities.¹³⁰ We might, *qua* noumena, not have any causal

the internal causes that determine the turnspit to act are mechanistic whereas the internal causes that determine us to act include our mental representations; since we want a conception of freedom that applies to human agents but not to turnspits, we can appeal to the fact that the latter seem to lack a distinctive kind of agency that the former have, and the presence of mental causation seems to account for this difference. Kant can respond in two ways. First, he can say that part of our sense that we unlike turnspits are free is due to our conviction that the grounds that determine our behavior are not wholly determined by past states over which we have no control at the time of action; there is no relevant difference in this respect between us and turnspits (and this accounts for a non-materialistic sense of 'mechanistic' according to which both a causally determined thinking being and a causally determined piece of matter is a mechanism). Secondly, Kant can offer an alternative version of his original argument that substitutes 'higher animal' for 'turnspit', and he can hold that whereas it seems to us that there is a difference *in type* between our behavior and that of higher animals that accounts for our sense of freedom, the compatibilist cannot make sense of this because on a compatibilist account there is only a difference in degree between the capacities to act on the basis of a representations of what is desirable that are exhibited by us and by higher animals. (Ameriks 2000, p. 241 raises a similar worry about the compatibilism Kant had adopted in the *Nova Dilucidatio*.) So Kant's anti-compatibilist argument in the Second Critique is not wholly dogmatic; it raises a genuine challenge for compatibilists to specify why mechanisms or organisms that we intuitively do not consider as free must count as free. But this cannot *suffice* as a critique of compatibilism: it has to be shown that the compatibilist interpretation of the notion that we have control over our actions are not satisfying; and it has to be shown, most of all, that there are elements in our self-conception as rational, responsible agents that strictly require Kant's ambitious notion of absolutely spontaneous self-determination.

¹³⁰ Hogan 2009 treats (a) Kant's commitment to the idea that our noumenal constitution lacks the kind of determining ground required for theoretically cognizing our noumenal self as equivalent to (b) Kant's affirmation of our noumenal freedom. But here he conflates (a) the idea that our only cognitively significant notion of a

capacities at all: Kant says that with regards to animals and inanimate matter – which, like all natural objects, can be considered both as appearances and as things in themselves – we have *no reason* for thinking that such objects have capacities that are not sensibly conditioned (CPR, B 575/A 547). So, the question is: why should we believe that we, unlike animals or turnspits, are free? How can we affirm the Rationality Thesis in the light of the No Experience Thesis, which states that there cannot be empirical evidence for our belief in our freedom? That is: *on what basis* can we consider ourselves free if freedom is not an object of experience?

In sum, Kant's theory of freedom, as reconstructed in this chapter, raises two main questions. First, are there conclusive reasons to prefer his ambitious concept of freedom over metaphysically modest compatibilist alternatives? Secondly, do we have any positive reasons to believe that we are free in this ambitious sense? If both of these questions could be answered affirmatively, Kant's conception might turn out to be uniquely attractive. Its *promise* lies in the combination of the naturalistic doubt about the adequacy of the libertarian attempt to explain our actions (for instance, by appeal to a notion of agent or substance causation) and the libertarian conviction that we can consider ourselves free in a metaphysically thick sense that cannot be simulated by 'shallow' naturalistic, compatibilist alternatives. In other words, Kant's position promises to synthesize (i) the intuition that the libertarian account of freedom is inadequate for theoretical purposes with (ii) the intuition that the naturalistic (compatibilist) model of freedom is inadequate for practical purposes and (iii) our sense that we can legitimately consider ourselves as free, responsible, norm-governed agents.

Whether Kant's theory can live up to this promise remains to be seen.

determining ground (as involving the schematized notion of causality) is inapplicable to things in themselves with (b) the idea that we are, qua noumena, capable of free action. Kant treats (a) as a merely negative point about things in themselves that follows from the fact that they are not in time (see CPR, B 149, and section 1 of my chapter I). The positive thesis (b) is not implied by this negative point. If Kant were to simply infer (b) from (a), he would be guilty of Beck's charge (1960, p. 190) that Kant is a 'panlibertarian' who believes that all objects (even chairs and donkeys) are free. That Kant rejects this inference is clear from CPR, B 575/A 547.

Chapter III: Transcendental Freedom and Practical Freedom

Kant's concepts of transcendental and practical freedom are closely aligned in all of Kant's critical writings, but there appears to be one major exception: namely, Kant's discussion in the Canon of Pure Reason. Here, Kant seems to employ a compatibilist or naturalistic conception of practical freedom which is strikingly at odds with his uncompromisingly incompatibilist discussion of freedom in the Transcendental Dialectic. In this chapter, I argue for an interpretation of the Canon that avoids the compatibilist reading and thus preserves the unity of Kant's theory of freedom in the First Critique. In the appendix, I explain my understanding of the many conceptual distinctions Kant draws in his discussion of freedom.

1: Transcendental and Practical Freedom in the Transcendental Dialectic

In the previous chapter, I argued that Kant is an incompatibilist about transcendental freedom: we are transcendently free only if we can act independently of determination by natural causes. In the Dialectic, Kant also introduces a closely related concept of *practical* freedom:

...the practical concept of freedom is based on this transcendental idea, and that in the latter lies the real source of the difficulty by which the question of the possibility of freedom has always been beset. (...) Obviously, if all causality in the sensible world were mere nature, every event would be determined by another in time, in accordance with necessary laws. Appearances, in determining the will, would have in the actions of the will their natural effects, and would render the actions necessary. The annihilation of transcendental freedom must, therefore, involve the elimination of all practical freedom. For practical freedom presupposes that although something has not happened, it ought to have happened, and that its cause, [as found] in the [field of] appearances, is not, therefore, so determining that it excludes a causality of our will – a causality which, independently of those natural causes, and even contrary to their force and influence, can produce something... and which can therefore begin a series of events entirely of itself. (CPR, A 534-5/B 562-3)

If there were no transcendental freedom, there would be no practical freedom either. This is because practical freedom is to be regarded as a kind of causality that operates independently of, "and even contrary to", the deterministic influence of natural causes. Surely, then, the concept of practical freedom delineated here is an incompatibilist concept: it inherits its incompatibilist connotations from the concept of transcendental freedom.¹³¹

But for all their similarities, the two concepts cannot simply be identified, not even in the Dialectic.¹³² "The transcendental idea of freedom does not by any means constitute the whole content of the psychological concept of that name, which is mainly empirical" (CPR, A 448-50/B 476-78). The 'psychological' concept of freedom is the concept of practical freedom.¹³³ I want to

¹³¹ Allison claims (1990, p. 57; see also Bojanowski 2006, pp. 201-2) that "in the Dialectic Kant is asserting a 'conceptual' rather than an 'ontological' dependence of practical on transcendental freedom": we need not *be* transcendently free in order to *be* practically free. I find it utterly implausible to read this deflation into Kant's claim that if transcendental freedom were *aufgehoben*, practical freedom would be *vertilgt*. See footnote 137 below for further discussion; see also chapter II for a detailed critique of Allison's deflationary interpretation.

¹³² This identification seems to me implicit in the discussions of Pereboom 2005 and Schönecker 2005.

¹³³ In the Critique of Practical Reason (KpV, 5: 96) Kant refers to 'psychological freedom' as well, but this cannot be the 'psychological' concept he has in mind in the Dialectic. This is because the psychological concept of freedom in the Second Critique is the one used by Kant's compatibilist opponents (which Kant himself rejects), who think that freedom is compatible with determinism as long as the acting cause is an inner, psychological cause; whereas the content of the psychological concept that Kant alludes to in the Dialectic is partly constituted by the idea of transcendental freedom, i.e., by the notion of independence from causal determination.

suggest that practical freedom relates to transcendental freedom as a more determinate to a less determinate concept. We can consider practical freedom as a specification of the more general notion of transcendental freedom because the idea of practical freedom provides a specification of the general features of transcendental freedom. Transcendental freedom is conceived as a cause whose causality is not determined by natural causes. This leaves open, first, the kind of causality that is being exercised, and, secondly, the kinds of determining causes from which this causality is supposed to be independent. The idea of practical freedom is more specific than that of transcendental freedom in precisely these two respects: it states that the kind of causality that is exercised through practical freedom is the causality of a *will*, and it specifies that this kind of causality must be conceived as operating independently from the necessitating influence of psychological, *sensuous* causes: "Freedom in the practical sense is the will's independence of necessitation through sensuous impulses" (CPR, A 534/B 562).

This explains why and in what sense practical freedom presupposes transcendental freedom: the specific properties exhibited by practical freedom can only exist if the general properties exhibited by transcendental freedom exist. There can only be the causality of a will whose exercise is not necessitated by sensuous impulses if there can be (generally speaking) a non-natural causality whose exercise is not necessitated by natural causes. This interpretation also allows us to make sense of Kant's suggestion that the concept of transcendental is a pure idea whereas the concept of practical freedom has empirical (psychological) content. Transcendental freedom is a pure transcendental idea whose content is independent of *all* experience (CPR, A 534/B 562). This idea forms part, but not "the whole" content of the concept of practical freedom (CPR, A 448-50/B 476-78). The concept of practical freedom contains the concept of an unconditioned causality (or that of an absolute spontaneity; see again A 448-50/B 476-78), but it also contains more, namely, a specification of the acting cause (in terms of the faculty of the *will*) and of the kinds of determining conditions (sensuous impulses) that must not necessitate this acting cause for it to be regarded as free. It is the appeal to these specific, *psychological* causes that provides the empirical content for the concept of practical freedom, since the concept of a sensuous impulse is empirical (it is drawn from 'inner' experience).

Since only part of the concept of practical freedom is drawn from experience, the concept as such cannot be considered an empirical concept. The fact that it has a 'pure' component – i.e., that the idea of spontaneity is part of its content – implies that it cannot be justified in the manner that is characteristic of empirical concepts, namely, by appeal to empirical evidence. That is, we cannot show on the basis of experience that we are capable of exercising our will independently of determination by sensuous causes.¹³⁴ But Kant is committed to the stronger notion that the concept of practical freedom cannot possibly be instantiated in experience, insofar as experience is understood as the objective representation of events¹³⁵: the exercise of a spontaneous causality is, as such, undetermined with regards to its temporal position (since it is not in turn caused by a temporally prior event), and having a determinate temporal position is a condition that any object of our experience must satisfy (CPR, A 532-3/B 560-1). Similarly, no observable act can be traced to an exercise of practical freedom as its cause within objective experience, because it is a condition on the objective experience of an event that it can be

¹³⁴ This is at least part of Kant's point when he argues that we cannot understand non-empirical causal powers on the basis of experience. See, for instance CPR, A 559/B 587, and KpV, 5: 46-7.

¹³⁵ See GMS: 4: 455; 458-9; and CPR, A 559/B 587.

referred to a cause from which it follows in a necessary, law-like manner (CPR, A 201/B 246-7); an exercise of free will cannot be considered as producing effects according to necessitating laws (CPR, A 534/B 562). Thus, there is a glaring gap between the idea of practical freedom and objective experience. This point will be important in section 2 below.

The proposal to view practical freedom as a 'species' of transcendental freedom requires an important clarification. One might think that this proposal implies that if the will is not determined by *sensuous* impulses but by some *other* (e.g., physical) kind of natural cause, the will is free. This seems like the natural consequence of putting a lot of weight on the idea that practical freedom is independence of determination by sensuous causes. But this implication must be rejected: all potential 'species' of transcendental freedom share the general feature of the concept which they render more determinate, namely, the feature of being independent of necessitation by *any* kind of natural causes. If the will is determined by any kind of natural cause, it follows that the will is neither transcendently free nor practically free. One might object that if just *any* kind of natural necessitation will simultaneously negate the general concept of transcendental freedom and all of its 'species', it is useless to delineate kinds of transcendental freedom by appeal to specific kinds of natural causes: if it does not matter to the issue of freedom versus determinism *which* types of natural causes determine the will, then why should we define practical freedom as the absence of necessitation by, specifically, sensuous impulses?

I think that for Kant it can be informative to focus on the absence of the necessitating influence of a specific kind of natural causes when delineating a specific kind of freedom, even if *every* kind of freedom requires independence from *any* kind of necessitating natural influence. This is so for two reasons. First, the definition of a kind of freedom by appeal to the absence of necessitation by a specific class of natural causes makes sense if the relevant philosophical issues revolve around the question of whether or not *these* causes determine our actions. In the case of practical freedom, the focus is on the faculty of the will. For Kant, the issues here concern the question of whether we can choose independently of the influence of *psychological causes* such as desires, emotions, inclinations, passions: these are the motive causes that are used in everyday psychological explanation of actions and that (Kant thinks) can at least in principle be cited in the prediction of actions¹³⁶; moreover, it is these sensuous motive causes whose influence threatens the purity of moral motivation.

Second, there is a very important systematic reason why Kant does not take into account the possibility that the will is determined by *non-psychological* natural causes. The point is not that we could call ourselves practically free if we knew that our will is exclusively governed by physical rather than psychological causality. If we knew this, we would have every reason to deny that we are practically free. The point is rather that the assertion that mental states and events (such as choices and acts of reasoning) are determined by physical forces requires a demanding epistemological, metaphysical and scientific framework that allows for the cognition of the influence of physical causes on phenomena of inner sense. For Kant, it is very doubtful that such a framework could exist: Kant accepts a kind of empirical dualism (CPR, A 379-80) according to which the empirical mind and body are wholly different empirical objects (appearances); in the same vein, he holds that natural causes that can be intuited *merely* through inner sense, such as psychological causes, are intrinsically different from natural causes that can

¹³⁶ See section chapter II, section 2 for a (qualified) defense of this point. Note that Kant does not think that it is epistemically possible for us to predict our actions.

also be intuited through outer sense, such as the forces examined by physical science. Correspondingly, Kant accepts that the causal laws governing the empirical mind or will are intrinsically different from the causal laws governing body (CPR, A 683/B 711). Given this view, it makes perfect sense to focus on psychological rather than physical causes in discussing whether the acts of the will are exclusively governed by necessitating natural laws.

One might raise a further question here. It seems that my suggestion that practical freedom is a kind of transcendental freedom presupposes that there is conceptual space for *other* kinds of transcendental freedom. A general concept might have only one species (or even none at all); but it seems as if it must be conceptually possible for a general concept to divide into distinct kinds. I agree: on my interpretation, the concept of transcendental freedom does not exclusively apply to the practical faculty of the will, but also to the theoretical faculty of the understanding. I discuss Kant's conception of freedom of thought in chapter IX.

2: Transcendental and Practical Freedom in the Canon of Pure Reason

The major obstacle to viewing practical freedom as a specific instance of the uncompromisingly incompatibilist type of transcendental freedom stems from the Canon of Pure Reason. In the extremely puzzling section concerning the final purpose of the use of reason, Kant says (CPR, A 802/B 830-A 804/B 832; my emphasis):

I must first remark that for the present I shall employ the concept of freedom in this practical sense only, leaving aside that other transcendental meaning which cannot be empirically made use of in explanation of appearances, but is itself a problem for reason... A will is purely animal (...) which cannot be determined save through sensuous impulses, that is, pathologically. A will which can be determined independently of sensuous impulses, and therefore through motives which are represented only by reason, is entitled free will (...) and everything which is bound up with this will, whether as ground or as consequence, is entitled practical. **Practical freedom can be proved through experience.** For the human will is not determined by that alone which stimulates, that is, immediately affects the senses; we have the power to overcome the impressions on our faculty of sensuous desire, by calling up representations of what, in a more indirect manner, is useful or injurious. But these considerations...as to what is good and useful, are based on reason. Reason therefore provides...objective laws of freedom, which tell us what ought to happen -- although perhaps it never does happen -- therein differing from laws of nature, which relate only to that which happens. (...) **Whether reason is not, in the actions through which it prescribes laws, itself again determined by other influences, and whether that which, in relation to sensuous impulses, is entitled freedom, may not, in relation to higher and more remote operating causes, be nature again, is a question which in the practical field does not concern us**, since we are demanding of reason nothing but the rule of conduct; it is a merely speculative question, which we can leave aside so long as we are considering what ought or ought not to be done. **While we thus through experience know practical freedom, as one [of or from: "von"] the causes in nature**, namely, to be a causality of reason in the determination of the will, transcendental freedom demands the independence of this reason...from all determining causes of the sensible world. Transcendental freedom is thus...contrary to the law of nature...and so remains a problem. But this problem does not come within the province of reason in its practical employment (...) The question of transcendental freedom is a matter for speculative knowledge only, and when we are dealing with the practical we can leave it aside as being an issue with which we have no concern.

What Kant says here seems to be strikingly at odds with what he says in the Dialectic. There, as we saw, he claimed that the idea of transcendental freedom is part of the concept of practical freedom, which implies that the concept of practical freedom cannot be shown to be instantiated in experience. But in the above passage, Kant seems to suggest (i) that we can assert that we are

practically free while bracketing the question of whether we are transcendently free, and (ii) that we can prove our practical freedom on the basis of experience. Many commentators respond to this seeming conflict by claiming that in the Canon Kant is operating with a compatibilist or naturalistic concept of practical freedom. The sharp conflict with his account in the Dialectic is explained by appeal to a 'patchwork thesis', according to which Kant's discussion in the Canon is influenced by remnants of the era of his earlier, pre-critical thought when he held a compatibilist view.¹³⁷ For reasons I will mention later, I accept that Kant's account in the Canon falls short of some of his considered critical views, but I also think that reading a naturalistic/compatibilist notion of practical freedom into the Canon is very problematic, for three reasons.

First, it is clear – and conceded by proponents of the naturalistic reading¹³⁸ – that Kant uses a non-naturalistic, incompatibilist concept of practical freedom *in the Canon*, and indeed only a few passages prior to the passage quoted above: "If the will be free, this can have a bearing only on the intelligible cause of our volition" (CPR, A 798/B 826). 'The intelligible cause' is to be contrasted with the empirical cause of our actions. It is surely uncharitable to suppose that Kant employed two wholly distinct notions of practical freedom in one section; and it is mysterious how he could completely switch from a non-naturalistic to a naturalistic notion of practical freedom within a few pages without noticing it.

Second, the idea (which is central to the naturalistic reading¹³⁹) that Kant's notion of practical freedom and practical reason in the Canon designates a natural cause of phenomena (i.e., as something that can be invoked in a theoretical explanation of why something happens), raises a major problem. Kant says, in the passage under dispute: "Reason therefore provides laws which are imperatives, that is, objective laws of freedom, which tell us what ought to happen...therein differing from laws of nature, which relate only to that which happens." So Kant here seems to be very much aware of the contrast between practical laws and natural laws that also looms large in the Dialectic – i.e., the categorical contrast between what *does* or *will* happen (or *why* things happen as they do) and what *ought* to happen. But if we suppose that he is operating with a naturalistic conception of practical reason as an 'Erklärungsgrund', this categorical contrast cannot be upheld. For on this supposition, when someone acts on reason's prescription that x ought to happen, we can cite reason's prescription that x ought to happen (or

¹³⁷ This view is defended, in different forms, by Carnois 1987, p. 29; Forschner 1974, p. 187; Gunkel 1989, p. 94-105; Kvist 1978, p. 84 ff.; Schönecker 2005, pp. 98-105; Steigleder 2002, pp. 4-8. The patchwork reading is challenged by Allison and Bojanowski. Allison claims (1990, pp. 60-4, and 1996, p. 109 ff.) that the key to reconciling Canon with Dialectic lies (a) in the idea that practical freedom in the Dialectic is only 'conceptually' dependent on transcendental freedom and (b) in the idea that practical freedom is independence from sensuous *necessitation* whereas transcendental freedom is independence from sensuous *affection*. Bojanowski's attempt to resist the patchwork reading differs from Allison's only insofar as he places (2006, pp. 201-2) the entire weight on (a). I think that (a) is manifestly implausible (see footnote 131 above). Allison finds support for (b) in some lecture notes, but the published texts themselves do not at all support the notion that transcendental freedom requires independence from affection by sensible causes. In the passage quoted above, Kant says that transcendental freedom "demands the independence...from all determining causes of the sensible world" (CPR, A 803/B 831). *Determining* causes are *necessitating* causes; and hence the intended contrast in the quoted Canon passage is between two forms of necessitation: first, in the case of practical freedom, necessitation by sensuous incentives; and secondly, in the case of transcendental freedom, necessitation by a class of natural causes that operate 'differently and more remotely' than sensuous incentives. Of course, the question is how this contrast is to be understood in the Canon; I discuss this issue below.

¹³⁸ See Schönecker 2005, pp. 117-18.

¹³⁹ See Schönecker 2005; Forschner 1974, p. 183 ff., and Kvist 1978, p. 88.

the agent's awareness of that prescription) *as explanatory factors in the chain of natural causes* and hence as parts of the content of a natural law that governs human action. The distinction between (normative) reason and (natural) cause that corresponds to Kant's contrast between 'ought to happen' and 'will happen' is then collapsed. One consequence of this collapse is that when an agent does *not* act on reason's prescription, her awareness of this prescription can still be cited as part of the content of the natural law that explains the agent's action, since it is one of the psychological factors to which the agent was subject at the time of action. (In a deterministic world, it is possible that the agent's awareness of a practical law might necessitate her to act contrary to the prescription.) But then the prescription is part of the deterministic causal nexus that explains why the agent, on a given occasion, *could not* have acted on the prescription. Reason's 'ought' would then have to be regarded as a psychological influence within the deterministic causal nexus that fails to ground a 'can'. This would imply that we can appeal to an ought-prescription both to prescribe how the agent should have acted and to explain why the agent could not have acted on this prescription.

The third point that tells against the claim that Kant is operating with a naturalistic conception of practical freedom in the Canon is that this claim is at odds both with what Kant announces to do and with what he actually goes on to do in the Canon. Immediately prior to those passages in which he is supposed to introduce the naturalistic conception of freedom, Kant announces that he is going to wholly leave aside everything empirical ("by leaving entirely aside any...empirical factors", CPR, A 801/B 829). It seems almost schizophrenic for Kant to follow up on this announcement by introducing an empirical, naturalistic notion of practical freedom. Moreover, it is clear that in his subsequent discussion Kant's needs are served by nothing less than an anti-naturalistic concept of practical reason and freedom. The general point of the Canon is to specify what is presupposed and required by the final end of the pure employment of reason. This final end is morality (CPR, A 801-2/B 829-30: "...the ultimate intention...has indeed...been directed to moral interests alone"). But for Kant, even in the Canon, what morality requires is

...that there really are pure moral laws which determine completely a priori (without regard to empirical motives, that is, to happiness) what is and is not to be done, that is, which determine the employment of the freedom of a rational being in general; and that these laws command in an absolute manner (not merely hypothetically, on the supposition of other empirical ends), and are therefore in every respect necessary. (CPR, A 807/B 835)

This requirement can be met only by an anti-naturalistic conception of practical freedom, because the notion of a *pure* moral law requires that the faculty of reason is, at least in its *prescriptive* function, independent of sensuous incentives ('empirical motives'). But if morality requires, with regards to its prescriptions, a non-naturalistic conception of practical reason, it would be odd and counterproductive for Kant to introduce, in a discussion aimed at investigating the requirements of morality, a naturalistic notion of practical freedom and reason.¹⁴⁰

These, then, are three compelling reasons to be skeptical about the idea that Kant uses a naturalistic or compatibilist concept of practical freedom in the Canon. But of course the passage quoted at the beginning of this section does yield prima facie support for the naturalistic reading. I now want to look more closely at this passage and see whether its putative naturalistic connotations might be defused, contrary to the compatibilist reading that is advanced, for

¹⁴⁰ Again, this is acknowledged by Schönecker 2005, p. 172: "Der [naturalisierte] Freiheitsbegriff spielt nach der Problempassage überhaupt keine Rolle mehr, der transzendente wird dagegen systematisch verwendet."

instance, by Dieter Schönecker (whom I consider the most careful advocate of this reading).

Let us focus first on Kant's claim that we can *prove* our practical freedom *on the basis of experience*. One strategy for defusing the compatibilist implications of this claim rests on the idea that it is not Kant's notion of practical freedom, but Kant's notion of 'proof' that should be understood in an aberrant sense, namely, as being tailored to the practical purposes that he goes on to mention (consider his appeal to a speculative "question which in the practical field does not concern us", CPR, A 803/B 831).¹⁴¹ This would require a correspondingly aberrant notion of experience, since Kant's notion of what can be proven and what can be experienced typically go hand in hand.¹⁴² And indeed, Kant introduces an 'odd' sense of experience in the Canon:

Pure reason...contains, not indeed in its speculative employment, but in that practical employment which is also moral, principles of the possibility of experience, namely, of such actions as, in accordance with moral precepts, might be met with in the history of mankind. For since reason commands that such actions should take place, it must be possible for them to take place. (CPR, A 807/B 835)

The idea that the concept of the moral law, and thereby the concept of a free and *pure* practical reason whose prescriptive acts legislate practical laws ("the actions through which it prescribes laws", CPR A 803/B 831), is presupposed by the 'experience' of moral actions is surely *very* different from the idea that (say) the concept of causality is presupposed by the experience of events in time. Let us bracket, for present purposes, the question of whether this line of thought has any plausibility. Our question is whether an appeal to an aberrant sense of proof and experience might help to defuse the compatibilist connotations of the claim that practical freedom can be proven on the basis of experience. Schönecker rejects this suggestion because he thinks that the naturalistic reading follows from the fact that Kant refers to practical freedom *as a natural cause*.¹⁴³ It is not clear whether Kant really deems practical freedom a natural cause¹⁴⁴, but even if he did, Schönecker's argument here faces the following problem: if it is granted that Kant may be operating with an aberrant sense of 'proof' or 'experience' according to which the purity and freedom of practical reason must be granted as a condition of the possibility of observable moral actions, then it seems but a very short step from there to the idea that he is also operating with an aberrant notion of a natural cause. In considering practical freedom a natural cause, he might not be intending the standard sense of natural causes as figuring in law-like explanations, but a permissive sense that merely requires that we can 'experience' (through the phenomenology of practical deliberation) practical reason as issuing prescriptions that are not conditioned by empirical motives and as overcoming the influence of sensuous impulses in choosing what to do. So even if Kant were advocating the claim that practical freedom is a natural cause, this need not imply the compatibilist reading.

¹⁴¹ See Allison 1990, p. 59 and Ameriks 2000, pp. 194-96 for a lucid formulation of this line of thought.

¹⁴² I owe this point to Funke 1981.

¹⁴³ See Schönecker 2005, p. 81.

¹⁴⁴ I am not convinced that Kant designates practical freedom as a natural cause. The relevant sentence reads, "Wir erkennen also die praktische Freiheit durch Erfahrung, als eine von den Naturursachen, nämlich eine Kausalität der Vernunft in Bestimmung des Willens..." My naïve reading of this sentence, undoubtedly fueled by my reliance on an incompatibilist conception of practical freedom, has always been that Kant means that we 'cognize' practical freedom not *as a natural cause*, but *as a freedom from natural causes* (als eine Freiheit von den Naturursachen). Schönecker admits that this reading makes rhetorical sense (2005, p. 81: "...grammatisch und rhetorisch...nicht unplausibel"), but rejects it because Kant goes on to draw a contrast between practical and transcendental freedom. This is the crux of the issue here, which I go on to discuss in the main text.

But I think that Schönecker has a stronger point here. The entire strategy of trying to defuse the seemingly compatibilist claims about practical freedom that Kant advances in the Canon by appeal to aberrant notions of 'experience', 'proof', or 'natural cause' faces the problem that Kant in the Canon clearly intends to separate his notion of practical freedom from the incompatibilist notion of transcendental freedom:

While we thus through experience know practical freedom to be ["eine von": *one of* OR *one from*] the causes in nature, namely, to be a causality of reason in the determination of the will, transcendental freedom demands the independence of this reason -- in respect of its causality, in beginning a series of appearances -- from all determining causes of the sensible world.

It seems that the best explanation for the fact that Kant draws this striking contrast between practical and transcendental freedom is that in the Canon, as opposed to the Dialectic, he is operating with a metaphysically modest, compatibilist concept of practical freedom.

I want to suggest an alternative explanation that draws on the difference between transcendental and practical freedom that Kant posits already in the Dialectic. We saw that in the Dialectic Kant explains practical freedom specifically as the will's independence from sensuous impulses; by contrast, his notion of transcendent freedom involves an appeal to the absence of determining natural causes *simpliciter*. I think that it is not implausible to read this conceptual distinction into the implied contrast in the Canon. Here, as in the Dialectic, practical freedom is understood as the freedom of the will from *sensuous* impulses. The above passage contrasts practical freedom thus understood with transcendental freedom understood as independence from *all determining causes of the world of the senses*. One must not confuse 'sensuous impulses' with 'determining causes of the sensible world' here. The determining causes of the sensible world are all the determining causes that we can posit in the phenomenal world, and this may include 'inner', sensuous causes such as desires, but it might also include 'outer', physical forces (e.g., gravitational attraction). So the contrast that Kant makes use of in the Canon might be that between practical freedom qua independence from 'inner', sensuous causes and transcendental freedom qua independence from all natural causes, including 'outer' ones.

Let us see whether this suggestion helps with the following problematic remark:

Whether reason is not, in the actions through which it prescribes laws, itself again determined by other influences, and whether that which, in relation to sensuous impulses, is entitled freedom [i.e., *practical* freedom], may not, in relation to higher and more remote operating causes, be nature again, is...a merely speculative question, which we can leave aside so long as we are considering what ought or ought not to be done. (CPR, A 803/B 831)

The question is what Kant has in mind when he appeals to 'higher and more remote operating causes'. Proponents of the naturalistic reading think that this phrase refers to sensuous incentives that operate on the will not directly (such as instincts) but mediately (via, say, reason's representations of states of affairs as painful or pleasant in the long run).¹⁴⁵ But this strikes me as implausible in view of the following remark (from the long passage quoted above on p. 54):

A will is purely animal (...) which cannot be determined save through sensuous impulses, that is, pathologically. A will which can be determined independently of sensuous impulses, and therefore

¹⁴⁵ See Schönecker 2005, p. 99; Beck 1963, p. 178, Paton 1958, p. 260 ff. This would suggest that practical freedom is freedom only from *direct* or *immediate* necessitation by sensuous incentives qua instincts, whereas transcendental freedom is freedom from all kinds of sensuous incentives, even those that operate only mediately.

through motives which are represented only by reason, is entitled free will....

Here Kant distinguishes the human from the animal will by saying that the latter "cannot be determined save through sensuous impulses" whereas the former can be determined by representations of reason "independently of sensuous impulses". If Kant thought that a human and an animal will differ only in degree, namely, in that the former is determined mediately by sensuous impulses (rather than immediately by instincts), he should have put his point in terms of *this* distinction (rather than explaining human freedom in terms of independence from sensuous impulses *simpliciter*); and he should have said that an animal will is determined by sensuous *instincts* rather than impulses.¹⁴⁶

I suggest that what Kant means by 'higher and more remote operating causes' is 'natural causes other than sensuous causes': in particular, *physical* forces. If so, we can read the above passage from the Canon in a manner that confirms, rather than contradicts, the view (introduced in the Dialectic) that transcendental freedom designates, with maximum generality, a causal capacity that is unconditioned by all types of natural causes, while the notion of practical freedom adds more specific (and partly empirical) content to this conception by specifying the freely acting faculty as a will and by specifying the relevant natural causes as sensuous impulses.

How does this suggestion bear on the interpretation of the problematic Canon passage? Kant emphasizes that for *practical purposes* we can assume that we are practically free and bracket, or leave out of account, the question of whether we are transcendently free. This need not imply that his concept of practical freedom here is a compatibilist concept. To say that we can ignore, for practical purposes, a certain possibility and venture a certain assumption is not to say that it would not matter to the truth of this assumption whether or not said possibility did obtain. For instance, it could be argued that we can ignore, for practical purposes, the possibility that we are brains in a vat and assume that we are moving in space, even though this assumption would be false if the possibility that we are ignoring did obtain. Whatever reasons we might have for ignoring the possibility that we might be brains in a vat, the possibility that we might be brains in a vat is incompatible with the truth of the assumption that we moving in space. Likewise, whatever reasons we might have for ignoring the possibility that there might be no transcendental freedom, it does not follow from the fact that we can ignore this possibility, and assume that we have practical freedom, that the assumption that we have practical freedom is compatible with the possibility that there is no transcendental freedom.

But this raises the following question: what is the point of bracketing, for practical purposes, the question of whether we are transcendently free, if we can 'prove' (in an aberrant

¹⁴⁶ The contrast between immediate and mediate determination by sensuous impulses may be at issue in Kant's early metaphysical lectures, but I find no evidence for the view that it is also in play in the Canon. One might think that Kant has this distinction in mind when he says: "the human will is not determined by that alone which stimulates, that is, immediately affects the senses". But the sentence continues: "we have the power to overcome the impressions on our faculty of sensuous desire, by calling up representations of what, in a more indirect manner, is useful or injurious. But these considerations, as to what is desirable in respect of our whole state, that is, as to what is good and useful, are based on reason." The emphasis here is squarely on the impact that *reason*, rather than *desire*, has on the will: for *any* influence of a sensuous impulse on the will, mediate or immediate, must count as resulting from 'an impression on our faculty of desire', and Kant here states that reason has the power of overcoming impressions of this kind *as such*. Likewise, every desire, as something that stimulates, can be traced to the immediate *affection* of an object on the senses, without this resulting in *determination* of the will either mediately or immediately.

sense) our practical freedom from sensuous impulses? Kant says that the question of whether we are free not just from sensuous impulses, but also from non-sensuous natural causes is "merely speculative" and need not concern us for practical purposes. If we suppose that Kant thinks that we can 'prove' our independence from sensuous impulses in an aberrant sense, it seems that we should expect Kant to say that we can also 'prove', in this aberrant sense, our independence from natural causes *other than* sensuous impulses. Why, on my interpretation, does Kant refuse to take that route and instead insists that the issue of whether we are determined by non-psychological natural causes should be left out of account (as a 'merely speculative' issue) altogether?

Two points are relevant here. First, it is likely that Kant's aberrant notions of 'experience' or 'proof' here are meant to relate to the fact that from our perspective of deliberating about what to do, it must seem to us that the course of our deliberation is not determined by the various desires or incentives to which we may be subject. For instance, I do not experience a desire to cheat on my partner as an inevitable force that leaves me with no option but to cheat on my partner. But if we accept that an appeal to this deliberative presumption (which is tied to the phenomenology of deliberation) 'proves', in an aberrant sense, that we can reason and choose without being determined by sensuous impulses, this has no implications whatsoever for the question of whether our deliberative course might not be determined by physical causes, i.e., by the influence of the body (brain) on the mind. While introspection may 'reveal' the sovereignty of our capacity for resisting the influence of sensuous impulses, we cannot even attach a clear sense to the notion of keeping at bay, in deliberating about what to do, the influence of physical forces. This is partly because the influence of sensuous causes, such as desires, on our deliberation and choice is a 'felt' influence; hence, it makes sense to say that we are conscious of resisting it, whereas the influence of physical causes on our mental life cannot become a felt object of self-conscious awareness. As far as the determination of our choice is concerned, physical causes operate too 'remotely' for the aberrant notion of our 'experience' of independence from determination by such causes to have any application.

Second, the question of whether we are determined by physical forces is 'speculative' in a special sense. The point is that the 'we' in question is our 'soul'; and so the question is whether aspects of our mental life, such as choices, can be determined by mechanistic forces, and how we could know whether or not mental states are determined in this way. This raises deep epistemic and scientific questions: first, about the possibility of interaction between mind ('soul') and body; and secondly, about how to cognize and to measure the effect of physical forces on psychological phenomena (phenomena that can only be given through inner sense, and that have no spatial location, extension, etc.). It is unclear how forces and powers whose effects can be systematically studied insofar as they relate to objects of outer sense are supposed to affect objects (events, processes, states) that can only be represented through inner sense.¹⁴⁷ It is on the basis of these kinds of considerations that Kant applauds a regulative principle for the psychological, empirical study of 'the soul' whose application ensures that "no empirical laws of

¹⁴⁷ The conceptual (and, in the technical sense, intuitive) framework for representing the effects of physical forces on 'outer' objects involves, to a significant extent, mathematical resources (i.e., the possibility of an a priori construction of the relevant forces and effects). Since these resources are either lacking altogether or extremely impoverished in the case of objects of mere inner sense, it is not clear what the basis for (conceptually and intuitively) representing the effects of physical forces on psychological states and events is supposed to be. This is part of the reason why Kant is profoundly skeptical about the scientific prospects of empirical psychology; see the Preface of the *Metaphysical Foundations of Natural Science* (and, for a brief discussion, my chapter VI).

bodily appearance, which are of a totally different kind, will intervene in the explanation of what belongs exclusively to inner sense" (CPR, A 683/B 711).

These two points, then, give us a sense in which the question of whether we are determined by non-sensuous natural causes (hence, whether we are transcendently free) remains speculative and intractable even if we are operating with a permissive, aberrant conception of what can be experienced and proven. The influence of physical causes on the will is 'too remote' to be taken seriously in anything but a decidedly speculative context.

In sum, I think that the appeal to the difference between physical and psychological natural causes can be used to make sense of the Canon passage cited above, to avoid the compatibilist interpretation of Kant's concept of practical freedom in the Canon, and hence to prevent a major clash between Canon and Dialectic. Since, as we saw, the compatibilist interpretation itself faces three major problems, my reading seems to me, on balance, satisfying from an interpretive standpoint, even though it is undeniably speculative in certain respects. We can add here that in his later writings (cf. the Critique of Practical Reason) Kant rejects the compatibilist conception of freedom of will as the 'freedom of a turnspit'.

Now, even though I think that the discussion of freedom in the Canon can be made to cohere with Kant's critical, incompatibilist conception of freedom, I find this discussion rather unsatisfying from a philosophical standpoint. We saw that Kant in the Canon relies on two major ideas: (1) An appeal to the 'experience' of deliberation 'proves', in a sense sufficient for practical purposes, that our choices are free from necessitation by sensuous causes. (2) An appeal to the 'experience' of moral actions 'proves' that practical reason is a pure, sensibly unconditioned source of necessary prescriptions, because this experience presupposes such prescriptions.

With regards to (1), the problem is this: even if we grant to Kant that the notion that choices might be determined by non-psychological natural causes raises intractably speculative issues, we are still left with the claim that our choices are free from determination by the states or properties that belong to our psychological constitution. If the relevant notion of freedom here is indeed irreducible to a compatibilist concept, it has very demanding metaphysical implications: it requires the genuine independence of the will from one domain of nature. This demanding, metaphysically loaded idea cannot be verified on the basis of our naïve, pre-reflective self-conception as deliberators. It might be responded that Kant's appeal to 'practical purposes' can be invoked to preempt both the metaphysical doubts about whether such freedom is possible and the epistemic doubts about whether we can know that we are free in this sense. But this line of thought is extremely problematic. It is obscure how I could take the fact that I am operating within a practical context to constitute a reason for believing that my will is free from psychological necessitation. The appeal to 'practical purposes' does not make the relevant metaphysical and epistemic problems go away; and it is, generally speaking, neither doxastically possible (barring a highly problematic kind of doxastic voluntarism which Kant rejects; see chapter IX) nor doxastically legitimate to form or hold onto a conviction by bluntly ignoring considerations that put pressure on this conviction.

(2) provokes the following complaint: the fact that we can observe actions that we tend to classify as morally right hardly requires the assumption that these acts are done for necessary moral reasons (they might, for all we can know, be based on contingent desires); and hence the observation of these acts does not presuppose the idea that our reason can issue necessary prescriptions independently of the influence of contingent sensible motive causes. So the

inference from our observation of actions we tend to classify as morally right to the idea that practical reason has a *pure* employment (that issues in necessary principles) is clearly fallacious.

Kant was clearly aware of these problems. In the Groundwork (GMS, 4: 459), we find him claiming that

freedom [of will] is a mere idea, the objective reality of which can in no wise be shown according to laws of nature, and consequently not in any possible experience; and for this reason it can never be comprehended or understood, because we cannot support it by any sort of example or analogy.

This – especially the rejection of the attempt to 'prove' freedom through examples or analogies – undercuts the notion that we may defend the idea that we have free will solely on the basis of our deliberative 'experience' of resisting a temptation. Moreover, we must notice that Kant's 'mature' ethical theory conflicts with the doctrine expounded in the Canon in at least one very important respect: namely, Kant comes to reject the claim, which he still endorses in the Canon, that we cannot comply with necessary moral laws without drawing on our desire for happiness (CPR, A 812-5/B 840-3). Once this claim is rejected, and the purity of moral motivation is emphasized, it follows that (pace (1)) neither the attempt to 'prove' the reality of such motivation by appeal to the phenomenology of practical deliberation nor (pace (2)) the attempt to prove the existence of pure moral laws by appeal to observable actions have any prospects of success. For all we can observe and experience, all moral laws might really be 'figments of the imagination', and no act might ever be based on necessary reasons (see GMS, 4: 407; 419).

So we would be well-advised not to place too much weight on Kant's discussion of freedom and morality in the Canon. It remains, of course, to be seen whether his later writings offer a more promising strategy to defend these concepts; I shall take up this question in my final chapter, where I examine Kant's attempt to defend the idea of freedom by appeal to our awareness of the moral law (the 'fact of reason'). The purpose of this section was mainly to solidify my interpretive claim that Kant's concept of practical freedom in the critical writings is an uncompromisingly incompatibilist notion.

3 (Appendix): Terminological issues concerning Kant's theory of freedom

The distinction between transcendental and practical freedom does not exhaust the relevant terminology that Kant employs in his critical philosophy. In this appendix I provide my understanding of his various concepts of freedom and of how they relate to each other.

First, I claim that the concept of *transcendental freedom* is to be used interchangeably with the concept of *spontaneity*. Kant introduces the very idea transcendental freedom by appealing to the idea of spontaneity: "the transcendental idea of freedom... stands only for the absolute spontaneity of an action" (CPR, A 448/B 476; compare A 448/B 476).

Next, consider Kant's distinction between *positive and negative freedom*. The negative concept of transcendental freedom designates, in its most general sense, independence from causal necessitation or determination (CPR, A 558/B 586). Given my account of practical freedom as a specific kind of transcendental freedom, there is also a specific version of the negative idea of transcendental freedom: as applied to practical freedom, this idea designates the will's independence from natural (psychological) necessitation (GMS, 4: 446; MS, 6: 213). These ideas are 'negative' insofar as they do not involve a positive specification of the laws according to which a causality that is independent from causal necessitation operates. As far as the will is concerned, the only laws of this type are moral laws, and hence the *positive* concept of freedom,

which Kant equates with *autonomy*, only applies to a will under moral laws (GMS, 4: 446-7).

Since both negative and positive freedom of will fall under the concept of transcendental freedom, and since the latter concept contains the idea of spontaneity, both the negative and the positive concept of freedom designate a form of spontaneity.

These points are, I think, rather uncontroversial; but things get very complicated when we take into account Kant's insight that the will of finitely rational beings divides into conceptually separable sub-capacities, one that is responsible for law-giving (will qua pure practical reason) and one that is responsible for volitional structure, that is, for choosing in accordance with or contrary to the laws of reason (Willkür).¹⁴⁸ The question is whether we can apply the concept of transcendental freedom or spontaneity to both practical reason and Willkür.

Here we must first notice a fundamental ambiguity in Kant's concept of *spontaneous causality*. When Kant introduces the notion of transcendental freedom, he explains it as a capacity to begin a state of one's own or to begin to act on one's own (CPR, A 533/B 561). This gives us a *generic* sense of spontaneous causality which corresponds to a wide sense of 'free agency', and which applies whenever we can speak of some effect resulting from the empirically unconditioned exercise of a capacity. However, when Kant applies the concept of transcendental freedom to the faculty of Willkür (CPR, A 534/B 562), his notion of spontaneous causality becomes more narrow, in two respects: first, it relates to the production of a *chain* or *series* of effects, and secondly it relates to effects that can be presumed or expected to occur *in experience* (CPR, A 537/B 565). This narrow sense of spontaneous causality is surely intended to track a notion of agency that relates to an *act of choice* that produces an observable effect, an 'outer' action which in turn has physical effects (so that we can speak of a chain of observable effects).

This distinction is at work in an extremely important passage in the Dialectic (CPR, A 548/B 576). In this passage, Kant claims (i) that natural grounds cannot produce a necessary ought judgment and that practical reason literally 'makes', with absolute spontaneity, an 'order of ideas' by declaring certain actions necessary. He adds (ii) that we presuppose that reason 'can have causality' with regards to actions that it declares necessary and, that reason expects *effects* from its ideas *in experience*. In (i), Kant employs the generic, wide sense of spontaneous causality: namely, the sense in which reason spontaneously produces representations or judgments. In (ii), Kant is concerned with causality and action in the narrow sense; he points out that we suppose that acts can occur in the world of experience due to the influence that the ideas (imperatives) pronounced by pure reason have on our choices (i.e., on Willkür). It is noteworthy that the narrow sense of free agency and spontaneous causality is *very* narrow indeed: Kant thinks of the chain of effects that he wants to ascribe to the influence of pure reason as comprising only *observable, outer* acts. When Kant says (CPR, A 548/B 576) that practical reason expects effects of its pure ideas in experience, he equates 'experience' with *outer* appearances, and he equates 'action' with 'outer action': he says that "the action to which the 'ought' applies must... be possible under natural conditions. These conditions, however, do not play any part in determining the Willkür itself, but only in determining the effect and its consequences [Erfolg] in the [field of] appearance." Here the determination of Willkür (e.g., the formation of a maxim) is explicitly contrasted with what occurs in the field of appearances.

¹⁴⁸ MS, 6: 226: "Laws arise from the will, viewed generally as practical reason; maxims spring from Willkür." Allison 1990, p. 130 ff. and Beck 1960, p. 190 emphasize that will (in the narrow, legislative sense) and Willkür are two aspects of one and the same faculty ('will' in a wide sense); as Allison puts it, they serve "a duality of function within the will".

These remarks are based on Kant's conviction that the only *perceivable* effect that can in principle be attributed to the influence of pure reason (although we can never know whether it really is due to such influence) is an outer action that occurs in space.¹⁴⁹ Kant's narrow sense of spontaneous practical causality plays such a prominent role in his exposition since it is of primary importance in our ordinary practice: when we make choices or adopt maxims, it is outer actions that we intend to perform; and it is only outer, physical behavior that can become an object of experience and that hence allows us to get a concrete, visible grip on what the efficient causality of pure reason might achieve. (I elaborate on these issues in chapter VII.)

Keeping this ambiguity in the notion of a spontaneous causality in mind, let us return to the question of whether the will can be considered free in both of its functions, that is, both as a source of practical judgments and laws (qua practical reason) and as a source of outer acts (qua Willkür). I want to argue, somewhat controversially, that we should answer this question affirmatively. That is: we should *not* withhold the applications of the notion of free, spontaneous causality to practical reason in its production of ideas, judgments or representations; and we should not suppose that *only* the narrow sense of spontaneous causality that is intimately tied to the production of observable effects designates a free causality. This is so for two reasons. First, although I think that Kant is right in assigning a prominent role to our conception of spontaneous causality as a source of *observable* effects, he cannot hold that it is strictly essential to the notion of a purely spontaneous causality that it be able to produce such effects. For he must acknowledge that our free Willkür can be causally efficacious even if its causality does not suffice for a production of observable effects: it is very important for Kant that a person's recognition of the force of moral reasons may have causal efficacy in that she adopts a pure maxim even if she lacks the physical power and causal efficacy that would be required for realizing the object of her morally pure intention in the observable world.¹⁵⁰ Second, the causal implications of the notion that our practical reason *produces* representations ('oughts') are impossible to deflate (the German word, "hervorbringen", relates to an efficient causality); and it is perfectly intelligible to ask whether this production is based on an empirically conditioned or unconditioned causality, that is: whether a representation or judgment is produced by a contingent natural ground (a 'want') or rather (as Kant claims, CPR, A 548/B 576) by pure reason. Hence, both practical reason and Willkür can be regarded as transcendently free or spontaneous in the *negative* sense: both have a spontaneous causality in that they are capacities to produce an effect independently of natural necessitation.

Kant's concept of *action* is also extremely multi-faceted, and extremely broad: in the Second Analogy, the causally determined exercise of mechanistic physical powers is regarded as an action, even as falling under 'the' concept of action (CPR, A 204/B 250). With respect to *free* agency, we must notice that Kant considers, somewhat confusingly, *both* the exercise of the capacity for 'intelligible' (noumenal) causality *and* the effect of this exercise as 'actions'; this is implied by the notion (CPR, A 544/B 572) of "an original action, through which something

¹⁴⁹ See, for instance, GMS, 4: 407: "...when the question is of moral worth, it is not with the actions which we see that we are concerned, but with those inward principles of them which we do not see." 'The actions which we see' are the *physical* manifestations of maxims – these, unlike maxims, are objects of direct perception; and we can only (fallibly and inconclusively) infer to the existence of maxims on the basis of perceivable outer acts.

¹⁵⁰ Consider Kant's example of a morally good will that, "owing to...the niggardly provision of a step-motherly nature...should wholly lack power to accomplish its purpose" (GMS, 4: 394). Note that this is perfectly compatible with saying that practical reason *expects* the realization of the objects of morally pure intention.

happens". The 'original action' is the spontaneous exercise of freedom (say, of the will), which brings about, as its effect, an event ('something that happens'). This event may be the formation of a maxim or the production of an effect in the observable, spatial world – both of which can also be considered actions. Corresponding to the narrow notion of spontaneous causality that I discussed earlier, Kant also has a *narrow notion of action* that applies only to outer actions (such as carrying groceries for an old lady). This narrow notion of action is at issue in his claim that when pure reason deems 'actions' necessary, it expects effects from his ideas 'in experience': when reason prescribes that one ought to behave respectfully toward the elderly, it expects that this prescription can result in (appropriately motivated) outer acts such as carrying groceries for an old lady. But this narrow concept is not the only or 'real' sense of free agency.¹⁵¹

Notice here that the spontaneous agency of practical reason that produces valid moral judgments is *pure*, which means that it is free not merely from natural *necessitation* but also (and even) from *affection* by natural causes. If practical reason is affected, in its exercise, by a contingent desire or inclination, the judgment it produces must fall short of representing a necessary law (although the judgment so influenced may, as a concealed hypothetical imperative, *deceptively* represent a practical law; see GMS, 4: 419). By contrast, the free exercise of Willkür (in the formation of choice) requires independence from natural *necessitation*, but always takes place under the influence of sensible *affection*.¹⁵²

I argued that both practical reason (in the production of moral judgments) and Willkür (in the formation of choice) can be considered (transcendentally and practically) free in the negative sense that designates the absence of causal necessitation. But can they also both be regarded as

¹⁵¹ At *one* point (MS, 6: 226), Kant denies that practical reason in its legislative function (qua producer of moral laws) can be called free since it is not concerned with action. This remark is the basis of Beck's 1960, p. 172, 185 and Bojanowski's 2006, p. 242-4, p. 261 insistence that practical reason has causality only with respect to its power to determine Willkür (i.e., to bring about morally good choices) but not with regards to its judgments. But we can take Kant at his word here only if we supply the *narrow* rather than the generic sense of 'action' and 'spontaneous causality'. That Kant thinks that we *act freely* (in what I called the *generic* sense) when we exercise our capacity for practical judgment is perfectly clear from what he says in the Groundwork (GMS, 4: 448), where he applies the famous notion that we must act under the idea of freedom to practical reason "with respect to its judgments" (compare CPR, A 803/B 831: "...reason in these actions, through which it prescribes laws..."). Likewise, Kant must accept that the negative concept of freedom or spontaneity, as designating the independence from sensible causes, applies to reason insofar as it 'makes' practical laws. In the 'Vorarbeiten' to the Metaphysics of Morals (VAMS, 23), we find Kant wavering, at one point (248) denying and at another point affirming (383) that the legislative will is free in the negative sense. But he is undeniably and strictly committed to saying that the will qua practical reason is free. The freedom of practical reason consists (GMS, 4: 448) in its not "receiving a bias from any other quarter with respect to its judgements" and consequently in the fact that it can "regard itself as the author of its principles independent of foreign influences". This answers accurately to the negative conception of transcendental freedom as applies to the will qua practical reason. See my chapters IV and V.

¹⁵² MS, 6: 213: "The human Willkür is in fact affected by [sensuous] impulses or stimuli, but is not determined by them; and it is, therefore, not pure in itself when taken apart from the acquired habit of determination by reason." Compare GMS, 4: 457-8, where Kant distinguishes *pure* reason that gives the law independently of sensibility from the faculty responsible for choice of maxims whose spontaneity is affected by sensible inclination. It is worth spelling out why *any* exercise of Willkür is essentially shaped by the influence of sensibility. This influence is obvious in the case where Willkür grants motivational weight to a sensible inclination in its choice of maxims. Now, we *can* exercise our Willkür in such a way that our will is moved solely by the rational influence of practical laws, but even then the influence of pure practical reason is experienced by Willkür as a kind of *force*, given the presence of sensible motives that are subordinated to the weight of moral reasons. So even in this case the influence of sensible affection 'shapes' the spontaneous exercise of Willkür.

positively free (as autonomous)? It should be uncontroversial that our positive freedom or autonomy consists in our capacity to establish ('legislate') practical laws and to determine actions in accordance with these laws, so that we may choose morally pure maxims and (should our physical powers suffice) realize the objects of our pure intentions in the observable world. The important point here is that even though a morally bad choice (which defies the weight of moral reasons) can be considered free in the negative sense (i.e., as independent of natural necessitation), only a choice based on moral reasons can be considered as having a law-like normative character. This is because the only practical reasons that have a universal scope and that involve a strict necessity (in that they apply to every agent regardless of her contingent psychological constitution) are moral reasons; since law-likeness requires strict universality and necessity, only a choice based on moral reasons is a choice that stands under laws of freedom and that is truly intelligible from the normative standpoint of pure (practical) reason. For Kant, such intelligibility and law-likeness is the hallmark of a positive conception of free agency. Only agency under moral laws can be considered autonomous, for only such agency is regulated by genuine laws that have their source in practical reason itself rather than in our contingent psychological constitution.¹⁵³

¹⁵³ Does Kant think that the will as a whole or rather its sub-capacity of pure practical reason deserves to be called autonomous? It is beyond dispute (cf. Allison 1990; Beck 1960; Bojanowski 2006; Willaschek 1992) that Willkür's choice of a *bad* maxim or action is *not* positively free or autonomous: a positively free will stands under moral laws (GMS, 4: 446-7). [Willkür is free in the negative sense even if it acts contrary to the moral law; this shows, in my view, that the distinction between positive and negative freedom is sufficient to absolve Kant of the charge that on his theory evil actions are not free. For recent versions of this complaint, see Prauss 1983 and Schönecker 1999, pp. 188-95; here I agree with Bojanowski 2006, p. 257.] But this leaves three possibilities: does the positive concept of freedom/autonomy apply (1) only to the will qua legislator of laws (Beck 1960, pp. 199-200); or (2) to the will as such in the wide sense (as including both legislative will and Willkür; Allison 1990, p. 132); or (3) to the determination of Willkür by the legislative will (Bojanowski 2006, pp. 256-7)? Bojanowski argues (p. 244) that (3) is recommended by the idea that Kant thinks that an appeal to the moral law gives us a positive conception of practical reason's capacity to spontaneously cause an entire series of appearances; and this suggests that we are autonomous/positively free if, and only if, our conception of the moral law determines the Willkür. But the fact that Kant wants to give positive reality to the idea that we have a *capacity* to produce effects in the world of appearances under the guidance of moral laws implies only that the will is autonomous if it legislates a principle that *can* suffice to motivate us; often what Kant says about autonomy relates only to there being such a *capacity* (GMS, 4: 461). This leaves us with (1) and (2). Option (2) seems plausible, as Allison correctly notes, because only the will in the broad sense is, in accordance with Kant's sense of practical autonomy (KpV, 5: 33), a law to itself (it, *qua* sub-capacity of legislative will, gives a law to itself *qua* sub-capacity of Willkür). However, pace Allison and Silber (1963, p. 181), the textual support for (1) cannot be denied. Kant says (KpV, 5: 33) that freedom in the positive sense *is* the legislation of *pure practical reason* ('will' in the narrow sense). Moreover, he applies (EEUK, 20: 225) the concept of autonomy to practical reason *and to the understanding* as law-givers, and this strongly suggests that the concept of autonomy applies, first and foremost, to our legislative capacities. I do not think that anything of substance hinges on this issue. Proponents of (1) and (2) can agree that we are practically autonomous, or positively free, insofar as we can articulate laws whose influence may suffice to determine us to will and to perform 'outer' acts.

Part 2
**The grounds of Kant's incompatibilism about
practical freedom**

Chapter IV: The Meaning of Practical 'Ought'

Kant believes that if human agents were exclusively parts of the nature ('if all causality were mere nature'), judgments that employ the concept of 'ought' would be meaningless:

It is impossible that anything in nature ought to be other than what in all these time-relations it actually is. When we have the course of nature alone in view, 'ought' has no meaning whatsoever. It is just as absurd to ask what ought to happen in the natural world as to ask what properties a circle ought to have. All that we are justified in asking is: what happens in nature? What are the properties of a circle? (CPR, A 547/B 575)

This is a bold and suggestive claim; understanding and assessing it is crucial for understanding and assessing Kant's incompatibilism about practical freedom. To understand and assess Kant's idea that ought-prescriptions (for present purposes, we shall confine our attention to *practical* 'oughts') are meaningless with regard to 'anything in nature', we first need to get clear about Kant's conception of the meaning of practical ought judgments. This is what I attempt in sections 1 and 2 of this chapter. On this basis, I show in section 3 that Kant's incompatibilist position rests on two conceptually separate arguments.

1: The meaning of categorical practical 'oughts'

Kant distinguishes between two different kinds of practical judgments. I will consider categorical oughts in this section and hypothetical oughts in the following section.

Kant's conception of the *content* of categorical practical judgments (for brevity's sake: CPJ) rests on his view that what is being expressed by a CPJ is a kind of necessity "that is not to be found in the whole of nature" (CPR, A 547/B 575). The only necessity to be found in nature is the material (nomological) necessity that is expressed by the claim, 'when a stone is dropped, it *must* fall to the ground'. We can also say that a CPJ and a statement about the natural course of events represent two distinct kinds of *reasons*. In 'The stone must fall to the ground if dropped', one is committed to there being a reason that *explains why* the stone cannot do anything but fall if dropped. But when I say that I must keep a promise, the reason that is being represented here does not explain anything that happens. As Kant puts it (CPR, A 548/B 576), a CPJ "does not follow here the order of things as they present themselves in appearance, but frames for itself with perfect spontaneity an order of its own according to ideas". The 'order framed' by a CPJ is an order of practical rationality. In framing such an order one does not aim at representing a course of events that are connected through causal and temporal relations but at representing events in terms of a set of distinctively normative concepts. The notion that I *must* keep my promises can be articulated in terms of the idea that I am *required* to keep my promises, or that I am *forbidden* to break my promises, or that it is *not legitimate* for me to keep my promises. The fact that the content of CPJ can be articulated in those terms shows that such judgments represent practical laws – normative laws of practical reason.

The fact that CPJ represent normative laws has evaluative implications: a practical law represents an action as necessary and as required because it represents this action as having a supreme value, as *unconditionally good*.¹⁵⁴ The unconditional worth of a morally required act is,

¹⁵⁴ More precisely, what the practical law represents as good (and necessary) is not simply a certain observable act, such as keeping one's promise, but rather a certain volitional structure (roughly, striving toward keeping one's promise on the basis of one's respect for the moral law, rather than on the basis of, say, one's fear of bad results;

negatively speaking, a function of (i) its motivational independence from any contingent empirical desires and (ii) the complete lack of bearing that its empirical consequences have on its worth.¹⁵⁵ In a more positive vein, a morally required act owes its unconditional moral worth to the fact that the agent is moved to act by her awareness of necessary practical reasons, that is, by her recognition of the unconditional value of the dignity of persons. There is, in other words, a congruence between the unconditional worth of a morally good choice and the unconditional value (dignity, or humanity as an end in itself) to which this choice is responsive.¹⁵⁶

Notice that the content of CPJ, as Kant conceives of it, requires an objective standard of practical rationality that allows for the classification of actions as good or bad, and as required, prohibited, or legitimate independently of an appeal to people's subjective inclinations. There can only be a valid practical *law* of practical reason which says that I must keep my promises if *rationally speaking I have no option but* to keep my promises. If it were rationally optional for me to keep or break my promises, or if I might just as well not keep my promises without incurring any rational deficiency, this would invalidate the idea that I *must* keep my promises or with the idea of there being a genuine law of practical reason that prescribes keeping promises with strict, uncompromising authority.¹⁵⁷

This must suffice as an account of Kant's conception of the content of CPJ. But to account for the distinctive meaning of CPJ, we must also focus on the distinctive *formula*¹⁵⁸ in which such laws are represented. The 'must' pronounced by practical reason strikes beings like us in the 'shape' of an *ought*. Kant emphasizes (CPR, A 548/B 576) that the idea that an action ought to happen implies that the action may happen but "perhaps will never take place"; and he says:

The representation of an objective principle, in so far as it is obligatory [nötigend] for a will, is called a command (of reason), and the formula of the command is called an imperative. All imperatives are expressed by the word ought, and thereby indicate the relation of an objective law of reason to a will, which from its subjective constitution is not necessarily determined by it (an obligation). They say that something would be good to do or to forbear, but they say it to a will which does not always do a thing because it is conceived to be good to do it. (GMS, 4: 413)

see GMS, 4: 403: "Necessity of acting from pure respect for the practical law is what constitutes duty").

¹⁵⁵ For (i), see KpV, 5: 26: if a determining ground was "only subjectively valid and merely empirical, [it would] not possess the necessity which is conceived in every law, namely, an objective necessity arising from a priori reasons". For (ii), see GMS, 4: 394; 416: "This imperative is categorical. It concerns not the matter of the action, or its intended result, but its form and the principle of which it is itself a result; and what is essentially good in it consists in the mental disposition [Gesinnung], let the consequence be what it may."

¹⁵⁶ This congruence between the unconditional value of a morally good will and the unconditional value of the dignity of persons (the recognition of the latter being the determining ground of a morally good will) is rarely spelled out explicitly by Kant (he comes close to it at GMS, 4: 430-1), but it is an essential aspect of his moral philosophy. A morally good will, defined as a will acting independently of any contingent interest in some empirical object or state of affairs, can exist only if there is some unconditional value to which any rational being can respond independently of whatever her contingent interests may be (i.e., which is "necessarily an end for everyone"; GMS, 4: 428). Kant argues that the only candidate for this type of value is the dignity of persons. On my view, acting on the basis of one's recognition of this value is tantamount to acting on the basis of respect for the moral law (after all, the second formula of the categorical imperative represents the value of dignity).

¹⁵⁷ As Adam Cureton reminds me, there are complications here about how to apply this point to the case of what Kant calls imperfect duties, but I shall have to omit these complications for the sake of simplicity.

¹⁵⁸ I use the term 'formula' instead of 'form' because Kant himself uses the term 'Formel' when referring to the imperative shape in which practical laws are represented to beings like us; and because his term 'form' is opposed to the 'matter' of an action (see, e.g., GMS, 4: 416). The contrast I am concerned with here is between the 'formula' and the 'content' of a practical judgment.

'Ought' is the concept through which a reason is represented to a being that *can* act but which *may also fail* to act on this reason. This formula reflects the fact that we are *imperfectly* rational beings whose subjective motives to act do not always accord with objective reasons. It expresses the attempt to subordinate the contingent motives which exercise an influence on the agent's will to the influence of practical laws. This influence is experienced as a kind of constraint by an imperfectly rational will; and hence the rational force that is at work in the subordination of contingent motives to necessary laws and reasons is called 'Nötigung'.

If there were a perfect congruence between a being's subjective motives to act and the objective demands on how it rationally must act, there would be no place for the subordinating influence that is expressed by the imperative formula. Such a being would not experience the rational influence of practical laws as a kind of constraint or force. Hence it would be pointless to address such a being in terms of an *ought*:

A perfectly good will would therefore be equally subject objective laws (viz., laws of good), but could not be conceived as obliged [genötigt] thereby to act lawfully, because of itself from its subjective constitution it can only be determined by the conception of good. Therefore no imperatives hold for the Divine will, or in general for a holy will; ought is here out of place, because the volition is already of itself necessarily in unison with the law. Therefore imperatives are only formulae to express the relation of objective laws of all volition to the subjective imperfection of the will of this or that rational being, e.g., the human will. (GMS, 4: 413-4)

The difference between the concept (or formula) through which a practical law is represented to a perfectly and an imperfectly rational will corresponds to a difference in the relation between practical judgment and prediction. In the case of an imperfectly rational will, the representation of a practical law does not imply compliance with the law. If one says, of a human agent, that she ought to keep her promises, this can be interpreted as, 'She will keep her promises *to the extent that she is rational*', or 'If she were perfectly rational, she would keep her promises'. This is compatible with conceding that she may, even will break the promise. But when a perfectly rational being represents the notion that it must do x, this strictly implies that it will do x.

These points are important for understanding Kant's cryptic claim (CPR, A 547-8/B 575-6) that "'ought' expresses a possible action the ground of which cannot be anything but a mere concept; whereas in the case of a merely natural action the ground must always be an appearance". If we trace an action to a natural 'determining ground' (i.e., to an empirical cause) such as a desire, the action is to be classified as a 'natural action' that follows necessarily, in a nomological sense, from the presence of the desire (this is obviously an oversimplification, since other psychological and 'outer' factors will contribute to the causal nexus). But when we judge that something ought to happen, we do not judge that the action will occur necessarily. Rather, we imply that it is (*merely*) possible that it happens, due to the possibility that the agent is moved to act by her recognition of the validity of the necessary reasons thought in the ideas (concepts) of pure practical reason. These ideas are addressed, in terms of 'ought', to imperfectly rational beings whose actions do not always conform to their conception of what is required of them.

In sum, the distinctive meaning of CPJ depends on two conditions: first, those judgments specify, in their content, a practical necessity or law; second, this content is represented through a formula that is tailored to the fact that we are imperfectly rational beings.

2: The meaning of hypothetical practical 'oughts'

We must now address the very difficult question of how Kant conceives of the meaning of *hypothetical* practical judgments (HPJ). We can start with a passage from the Groundwork:

Since every practical law represents a possible action as good and, on this account, for a subject who is practically determinable by reason, necessary, all imperatives are formulae determining an action which is necessary according to the principle of a will good in some respects. If now the action is good only as a means to something else, then the imperative is hypothetical; if it is conceived as good in itself and consequently as being necessarily the principle of a will which of itself conforms to reason, then it is categorical. (GMS, 4V: 414)

This passage suggests that *all* practical imperatives represent a necessity and a type of value. An action can be deemed good and necessary either in an unconditional sense (in which case we have a CPJ), or in a conditional sense, as a means to something else. In the latter case the imperative that represents the action as necessary is a hypothetical imperative, or rather, one of two kinds of hypothetical imperatives: either an imperative of skill (relating to an arbitrary purpose) or an imperative of prudence (relating to a purpose that all finite rational beings adopt, namely, their own happiness).¹⁵⁹ This suggests that CPJ and HPJ share a common core; and this impression can be strengthened by appeal to a passage from the Dialectic, where Kant is concerned with 'oughts' that relate to 'all matters of conduct' (CPR, A 547-8/B 575-6):

That our reason has causality...is evident from the *imperatives* which in all matters of conduct we impose as rules upon our executive powers. (...) ...the 'ought' pronounced by reason confronts such willing with a limit and an end – nay more, forbids or authorizes it. **Whether what is willed be an object of mere sensibility (the pleasant) or of pure reason (the good), reason will not give way to any ground which is empirically given.** Reason does not follow here the order of things as they present themselves in appearance, but frames for itself with perfect spontaneity an order of its own according to ideas, to which it adapts the empirical conditions, and according to which it declares actions to be necessary, even although they have never taken place, and perhaps never will take place. (My emphasis)

'The pleasant' is a place-holder for things that we deliberately pursue and that lack *moral* value. An ought judgment directed at 'the pleasant' is a hypothetical imperative because it declares that an action is necessary *if* one pursues the end of obtaining pleasure.

The passages considered so far suggest that it is not only Kant's conception of CPJ but also his conception of HPJ that is distinctively *anti-Humean*, in the sense that Kant endorses the following ideas: (I) Practical Reason can motivate the will in its pursuit of non-moral value in a manner that is irreducible to the contribution of desire; (II) There are objective standards of non-moral goodness that practical reason can establish and represent.

However, this anti-Humean picture of the hypothetical imperative conflicts with other textual data. About imperatives of skill, Kant says (GMS, 4: 415): "Here there is no question whether the end is rational and good, but only what one must do in order to attain it." This implies that *practical* reason is completely uninvolved in the derivation of HPJ: a HPJ can be correct or valid independently of whether a course of action would realize anything of value, a question which is characteristically answered by practical reason. The correctness of a HPJ solely depends on the correctness of the *theoretical* standard that yields the content of a HPJ:

¹⁵⁹ I will ignore here the distinction between 'problematic' and 'assertoric-practical imperatives' (GMS, 4: 414) that Kant himself rejects in the First Introduction to the Critique of Judgment (EEUK, 20: 200).

namely, on whether the HPJ correctly represents a necessary means for bringing about a certain result (the 'must' of empirical, causal necessity). That HPJ draw on theoretical rather than practical standards is confirmed by Kant's claim that coming up with imperatives of skill is a task for the practical part of 'the sciences' (GMS, 4: 415). In the Introductions to the Critique of Judgment, Kant denies that providing for 'technical-practical' rules requires a practical branch of philosophy, which means that practical reason plays no part in establishing those rules (KU, 5: 172): they are established through theoretical cognition (of causal relations; EEUK, 20: 199-200).¹⁶⁰ Furthermore, Kant sometimes explains his often-repeated point that hypothetical imperatives are 'analytic' by making descriptive claims such as, 'a will who wholly wills an effect also wills the means required for realizing this effect' (GMS, 4: 417). This suggests that an agent who wills an end quite 'automatically' also wills what she conceives as the means required for realizing this end. If she doesn't will those means upon learning what they are, it follows that she does not wholly will the end. There is no room for *practical reason* to influence the will through the representation of HPJ: to account for non-moral willing, we need no resources apart from contingently given desires and the discoveries of causal means-end reasoning. This suggests that Kant accepts that practical reason's role in governing conduct is exhausted by what it contributes, as *pure* reason, to *moral* legislation and motivation. When it comes to the deliberative pursuit of non-moral value, Kant appears to endorse (NI) The ('lower') faculty of desire is solely responsible for the determination of the will in the pursuit of non-moral ends; and (NII) The only objective standards and necessities to be established with regards to the pursuit of non-moral ends belong to theoretical reason. (NI) and (NII) jointly imply a picture of non-moral willing according to which reason's only role is to serve the bidding of our contingent, empirically given desires or inclinations:

Then only, when reason of itself determines the will (not as the servant of the inclination), it is really a higher faculty of desire to which that which is pathologically determined is subordinate, and is really, and even specifically, distinct from the latter... Reason, with its practical law, determines the will immediately, not by means of an intervening feeling of pleasure or pain, not even of pleasure in the law itself, and it is only because it can, as pure reason, be practical, that it is possible for it to be legislative. (KpV, 5: 24-5)

One can rightfully ask in what sense the concept of willing (as opposed to mere desiring) could even meaningfully be applied to this picture of non-moral motivation – a worry that is corroborated by Kant's remark that it is only in cases of moral willing that the faculty of practical reason is 'really, and even specifically' distinct from the 'lower' faculty of desire.

Hence, what Kant says about HPJ suggests *both* the anti-Humean claims (I) and (II) *and* their Humean negations (NI) and (NII); and so his discussion of hypothetical imperatives poses a major interpretive challenge. Unlike other commentators¹⁶¹, I think that neither the Humean nor the anti-Humean strand in Kant's thinking about HPJ can be ignored or explained

¹⁶⁰ Pollok 2007, pp. 67-8 argues that Kant's relegation of hypothetical imperatives to theoretical philosophy is not characteristic of the Groundwork, but, as we saw, Kant associates at least imperatives of skill with 'the sciences' even here (GMS, 4: 415).

¹⁶¹ Commentators tend to focus rather one-sidedly either on the Humean strand of Kant's conception of HPJ (cf. Bojanowski 2006, p. 28, p. 35, p. 77; Herman 2001), in which case they prioritize Kant's dismissive treatment of HPJ in the Second or Third Critique, or on the anti-Humean strand (cf. Allison 1990, p. 88 ff.; Hill 1973; Pollok 2007), in which case they altogether ignore or dismiss Kant's treatment of HPJ in the Second and the Third Critique. I will indicate my disagreement with these interpretations below.

away. In this section I shall try to combine these two strands into a coherent picture.

I think that we can properly understand the peculiarities of Kant's conception of HPJ only if we proceed from a firm understanding of his conception of non-moral ends. Kant thinks that our pursuit of non-moral ends relates to *happiness*, which he identifies with the continuous contentment or pleasure one takes in one's life.¹⁶² Now, the orthodox way of interpreting Kant's conception of happiness says that on his view all our non-moral acting is intentionally directed toward our own pleasure.¹⁶³ This form of hedonism is not very attractive (even if we keep in mind that for Kant, moral motivation – not aimed at pleasure – relates to a great variety of ends, such as the development of one's own talents and character¹⁶⁴). I think that we can ascribe a more moderate and more plausible form of hedonism to Kant, which draws on the idea that anyone committed to maximizing her pleasure is well-advised to refrain from constantly aiming at such maximization; for it seems to be a psychological fact that we get most satisfaction from pursuing projects for their own sake.¹⁶⁵ Kant's theory is well-suited to accommodate this point on the basis of the distinction between the choice of general maxims, where he must deem the concern with one's own pleasure the *only* (non-moral) criterion, and specific intentions (often formed without explicit reflection), where the concern with one's own pleasure need not be in the foreground.¹⁶⁶ For example, one forms the general maxim to take an interest in one's friends because one believes that close friendships make one happy. This maxim is not only compatible with but, arguably, *requires* that in one's everyday interaction with friends the concern with one's own pleasure is not in the foreground: one's intention to call up a friend is based on one's concern for her well-being. This 'higher level hedonism' seems to me quite plausible; whether it fits all relevant phenomena is a question I cannot discuss here.¹⁶⁷

¹⁶² Kp5, V: 22: "Now, a rational being's consciousness of the pleasantness of life uninterruptedly accompanying his whole existence is happiness." KpV, 5: 25: "To be happy is necessarily the wish of every finite rational being... For we are not in possession...of satisfaction with our whole existence...because we have wants and these wants regard the matter of our desires, that is, something that is relative to a subjective feeling of pleasure or pain, which determines what we need in order to be satisfied with our condition."

¹⁶³ The orthodox interpretation is endorsed by Beck 1960, pp. 92-102; Darwall 1983, p. 174; Irwine 1984, pp. 39 ff.; Johnson 2005; Meerbote 1984a, pp. 66-7; Wood 1984, p. 83.

¹⁶⁴ Hence, Kant can avoid Nozick's (1977) attack on *universal* hedonism: for Kant, pure reason (morality) supplies us with ends and ideals that prohibit hooking ourselves up to Nozick's experience machine.

¹⁶⁵ See Railton 1984 for a very convincing exposition of this point.

¹⁶⁶ Herman's 2001, p. 135 distinction between complex and simple ends is, I think, intended to make a similar point. The orthodox view that for Kant *all* non-moral intention aims at pleasure is also rejected by Allison 1990, pp. 102-3 and Reath 1989, 2006. Reath once (1989) took himself to be aiming at a more radically non-hedonistic reading, but in a more recent appendix to the original paper (2006, p. 56) he clarifies that he seeks to attribute to Kant a "sophisticated hedonism" after all. I must confess that I find certain aspects of his account, such as the distinction between hedonistic *criteria of choice* and non-hedonistic *ends* (2006, p. 45; compare Allison 1990, p. 103) elusive. A criterion of choice must, it seems, reflect reasons for acting and hence the end for the sake of which an act is chosen. So I think that the distinction between choices of general projects and choices of specific courses of action, or between complex and simple ends, is more promising as a moderation of Kant's hedonism.

¹⁶⁷ I think that the idea that the explicit and exclusive concern with one's own pleasure pertains primarily to one's choice of *general* projects and ends has a firm basis in the text: it is undeniable that for Kant the concern with one's own happiness arises, as *the* competitor against moral considerations, first and foremost at the level of choosing general maxims (KpV, 5: 75; 81). Moreover, Kant's conception of non-moral benevolence as "love for people and affectionate sympathy" (KpV, 5: 82) implies that a benevolent agent aims, in her concrete benevolent acts, at *other* people's happiness rather than her own, for affectionate sympathy simply *is* an other-directed emotion. The major difficulty in attributing this picture to Kant is that he undoubtedly does claim that the desire

Now, for our purposes it is crucial to get clear about Kant's claim that realizing one's own happiness is a purpose that we adopt as a matter of natural necessity. His appeal to natural necessity may seem puzzling here, for setting oneself an end is not the kind of thing one can be causally necessitated to do on Kant's conception of practical agency.¹⁶⁸ We can partly demystify Kant's point by assuming that what he means is not that we always act for the sake of happiness, but that happiness is the only thing that we, as a matter of empirical necessity, *always* and *uninterruptedly* desire or long for (KpV, 5: 25)¹⁶⁹, even if we do not act on the basis of this desire at all (as when we act on the basis of moral reasons). As we shall see, there are other important distinctions to be drawn here, but for our present purposes one point must be emphasized: while practical reason is not the source of our concern with our own happiness, it nevertheless represents a state of affairs in which a finite rational being is happy as valuable, and an action that secures this state of affairs as *good*.¹⁷⁰ Kant thinks that we are morally required to help others to attain happiness, and he argues that the correspondence between 'being worthy of happiness' and 'being happy' is the highest good of reason (KpV, 5: 124-5). So he must think that for a dependent rational being, attaining happiness – having its one ardent longing satisfied – is a valuable thing. What he denies is that it can be regarded as *unconditionally* valuable: the goodness of an act that secures a person's happiness depends on its moral permissibility. An impartial rational observer would approve of the uninterrupted happiness of a dependent being *provided that* this being shows traces of a good will (GMS, 4: 393). So whether or not an act that promotes my happiness lives up to moral standards is a condition on its goodness, or on my having any reason to perform it, whereas its conduciveness to happiness is not a condition on the moral goodness of the act.¹⁷¹ The idea that the moral impermissibility of an act *completely cancels, rather than merely outweighs* any reason for performing the act that may derive from its conduciveness to my happiness is implied by the idea that the value of my own happiness cannot be rationally weighed against the moral value of the dignity of persons.¹⁷²

Now, one reason why the fact that we all continuously desire our happiness as a matter

for an object is based on the anticipation of pleasure that the object would give one (KpV, 5: 23); this point is well-put by Johnson 2005, pp. 53-5. Pace Reath 2006, pp. 37 ff., Kant's point here clearly pertains not just to the causal genesis but (also) to the intentional structure of desire. One might, however, try to moderate Kant's claim by supposing that the hierarchical relation between general maxims and specific intentions is echoed in the relevant desires: while the desire to call my up my friend, or the desire to help an old lady with the groceries, need not be based on an anticipation of the pleasure I would get from these specific acts, these desires derive their motivational force from a higher-order desire (incorporated into my general maxim) that is based on the expectation that engaging in the relevant types of acts (without reflecting on and calculating the pleasure they might or might not bring in every instance) will make life pleasant overall.

¹⁶⁸ This may be why Herman (2001, p. 133) denies that Kant has natural necessity in mind here. But the textual evidence is clear (GMS, 4: 415): "There is...one purpose which [finite rational beings] not merely may have, but which we may with certainty assume that they all actually have by a natural necessity, and this is happiness."

¹⁶⁹ This distinction is, I think, similar in spirit to Hill's 1973, p. 437 distinction between 'having happiness as an end' and 'willing happiness as an end'.

¹⁷⁰ See KpV, 5: 62. Kant's term for the kind of value that reason attaches to the happiness of a dependent being is 'Wohl' ('welfare') rather than good; this is an empirical concept rather than a concept of pure reason, but practical reason nonetheless applies its sui generis concept of goodness to acts that secure an agent's happiness.

¹⁷¹ For the idea that morality is the "highest condition" of our "private purposes", see GMS, 4: 396; KpV, 5: 61-2.

¹⁷² This is the upshot of Kant's point (GMS, 4: 434-5) that the objects of human inclinations have a 'market value' that can be weighed against other such objects, whereas the dignity of persons allows for no equivalent and is elevated above all market value ("über allen Preis erhaben").

of natural necessity does not imply that we are necessitated to set ourselves any specific end is that happiness as such is an indeterminate or 'formal' purpose: for it to become an object of an agent's deliberative pursuit, the agent must fill it with empirical content. This means that a sensibly affected agent must figure out which concrete things would make her happy. It is crucial to notice that Kant holds that the rules concerning what concrete ends one should adopt are established on the basis of theoretical rather than practical reasoning.¹⁷³ Equally importantly, Kant thinks that these rules *lack necessity* and hence fail to be *laws*. He has three distinct arguments for this claim. First (1), he focuses on the equation of necessity with universality: rules of prudence could only be necessary if they applied to *all* of us. However, different people are subject to different rules of prudence, depending on their psychological constitution. Even with respect to one person, there is considerable flux in her given desires and hence in her conception of happiness over time (KpV, 5: 25). Thus, rules of prudence cannot apply universally and hence lack necessity. Second (2), Kant focuses on the equation of necessity and a priority: the necessity of a rule requires that we know it with certainty, but (GMS, 4: 417-8) in order to know with certainty whether realizing an end would make one happy one would have to be omniscient. Kant's basic idea here is correct: our psychological constitution is fluid in ways that we cannot safely predict, and thus our actions might not have the intended psychological effects¹⁷⁴; and unforeseeable circumstances (e.g., a sickness, a new love, etc.) may interfere with the psychological effects that our acts were intended to have.¹⁷⁵ We *do* have inductive grounds for supposing that, on average, some actions are more conducive to happiness than others, but this inductive basis supports only 'counsels' that fall short of epistemic certainty and hence necessity. Finally, notice that the conclusion that prudence cannot ground practical necessities does not depend on (1) and (2): it is already implied by (3) the idea that the value promoted by rules of prudence is only conditional and strictly subordinate to moral value.¹⁷⁶

A further point of crucial importance is that imperatives of skill are practically subordinate to counsels of prudence. By themselves, imperatives of skill relate only to a *possible*

¹⁷³ EEKU, 20: 196-7: "If precepts for the promotion of one's happiness are given...then all that is represented are the inner conditions of the possibility of such happiness...as an immediate consequence from the theory of the object in relation to the theory of our own nature (ourselves as a cause)..." Compare KU, 5: 172. Pace Reath 2006, pp. 45-6, I do not think that the *imagination* has a legitimate role in this process; its playing a part here would interfere with the standards of causal reasoning, by distorting the real psychological effects of causes and by diverting attention from inductive grounds that support the expectation of pleasure.

¹⁷⁴ Kant could support this point by appealing to his skepticism about empirical psychology (see my chapter II, section 2). In the Second Critique (KpV, 5: 26) he takes more drastic measures and argues that *no* predictive enquiry based solely on experience can ground anything but inductive rules that lack a priori certainty.

¹⁷⁵ Kant has a tendency to argue for this point by appeal to a (I think) rather artificial conception of happiness, namely (GMS, 4: 417), that of "a total sum, a maximum of happiness in my present and in every future state". Omniscience would be required to know whether an action could produce this "totality of an infinite number of consequences". It is not obvious that we ever aim for this ideal: our pursuit of happiness is compatible with a willingness to accept trade-offs, such as a short period of pain for the sake of a longer period of pleasure. After all, Kant himself argues that the temporary absence of pleasure supports the capacity to feel contentment in the long run; (Anth, 7: 231-2). But even if our actual pursuit of happiness is directed at happiness in the realistic rather than the exalted sense, Kant is surely right in emphasizing that we cannot know with any degree of certainty whether some choice would bring us the greatest possible net sum of happiness.

¹⁷⁶ See GMS, 4: 420: even if we could cognize certain means for obtaining happiness, it would be rationally legitimate to give up the end of happiness in favor of moral ends, whereas it is never rationally legitimate to give up moral ends in favor of one's own happiness. Hence, only the rules of morality are necessary laws.

end: they are practically irrelevant (that is, irrelevant to the willing of a rational agent) unless they concern a concrete end that is prescribed by a counsel of prudence.¹⁷⁷ This follows from the fact that (on Kant's view) our only non-moral reasons for acting derive from the ideal of happiness that is the object of counsels of prudence; so whatever imperative of skill we attend to for the sake of whatever non-moral end, our practical interest in this imperative derives from and depends on our interest in our happiness. Suppose, for instance, that I adopt the end of building a hut in my garden because I am guided by a counsel of prudence that states that sitting in the hut and watching the sunset will be enjoyable for me. I consult the technical rules for building a hut provided by the relevant kind of craft (or, as Kant sometimes puts it, by the concept of a hut); and I realize that in order to build a hut, I must first go and buy a certain type of wood. The imperative of skill, 'I ought to buy wood' could be understood as a short-hand clause for 'I must buy wood in order to build a hut'; but given the fact that the end of adopting a hut is based on my conviction that building a hut would make me happy, its complete content – *the kind of content that actually answers to the deliberative structure of my subjective principle of willing* – would be: 'I must buy wood in order to build a hut, and hence in order to become happy (to enjoy the sunset, etc.)'. The clause, 'in order to become happy' expresses – for the reasons discussed above – an inductively grounded probability (rather than a strict necessity).

The account of HPJ given so far can be understood as saying that we use theoretical reason for investigating the means to satisfy our empirically given desire for happiness. A point for *practical reason's* having a place in this picture can be made by noting that Kant's picture of non-moral willing leaves room for two kinds of *practical errors*. When we render our given desire for happiness more determinate by adopting a concrete end, it would be self-defeating *not* to go by known inductive (if inconclusive) standards of happiness. We all know the phenomenon of adopting an end on the basis of a mere whim evoked by something that appeals to one's imagination, and whose grip on us is so strong that it overturns the warning voice provided by counsels of prudence. For instance, suppose I adopt the end of building a hut under the influence of a TV commercial picturing a happy old man who watches the sunset from his garden hut, which leads me to ignore the counsel that draws attention to my inefficiency as a carpenter and my tendency to quickly grow tired of projects of this type.¹⁷⁸ Moreover, at the level of the deliberative pursuit of happiness where we try to realize concrete goals, we all know the phenomenon of not following through with our plans under the influence of stress, exhaustion, etc. Suppose I have adopted the end of building a hut on a prudent, informed basis. I have worked hard for half a year to build the hut, and it will take me two more weeks to finish up the work. I read that a big storm is coming up, and I know that in order to save what I have built so far, I need to cover the structures before the storm comes in. However, when I get home I dread the prospect of covering the structures; the influence of exhaustion inclines me not to take the necessary means for realizing my end but to indulge in mindless entertainment (e.g., cartoons).

¹⁷⁷ Kant's acknowledges this point at KU, 5: 172, where he sets apart "rules of art and skill generally" from rules of prudence "as a skill in exercising an influence over men and their wills".

¹⁷⁸ This kind of *weakness of will* that leads me to overturn a rule of prudence of which I am aware must be distinguished from a different type of mistake that is *wholly theoretical*: namely, that of adopting one's concrete ends on the basis of counsels of prudence that one believes to be true but that are in fact false (in the sense that they do not correctly represent a likely means for obtaining happiness). There is room for an interesting kind of mixed case here: namely, a kind of self-delusion or wishful thinking that occurs when one dismisses a correct rule of prudence of which one is consciously aware as false under the influence of desire.

It is instructive to spell out the implications that the thoroughly Humean picture has for these cases. On this picture, the will is nothing over and above a manifold of specific desires of varying strength. If I overcome my exhaustion and cover the structure, the Humean picture says that my desire for building the hut is stronger than my desire for rest and mindless entertainment. The problem is that if I give in to my exhaustion, the Humean picture forces us to draw the analogous conclusion that my desire for mindless entertainment is stronger than my desire for building the hut (unless it can be said that my theoretical reasoning is somehow faulty). On the Humean picture, whatever desire actually determines me to act expresses my will, and there is no room for a contrast between what I am actually motivated to do and what my real ends are, or for a practical criticism of my choices.¹⁷⁹ Most importantly, there is no room on this picture for the intentional (knowing) violation of a HPJ. If I adopt my concrete ends on the basis of a whim, and if counsels of prudence cannot influence me to choose otherwise, it follows that at that time I simply do not desire the goal (happiness, qua continuous contentment with my life as a whole) toward which these counsels represent means, or that I have a stronger desire for some other goal (short-term satisfaction); hence, I am not violating the counsels concerned with this goal. And whenever I do not take the means I know to be conducive to a concrete end 'e', it follows that I do not desire e (anymore), or that I have a stronger desire for something else; and then I do not violate a principle which says that one must take means that are necessary for realizing e.

This thoroughly Humean conception of non-moral willing cannot be attributed to Kant as his considered view. Kant explicitly affirms that *all* imperatives express a kind of *Nötigung* of the will (GMS, 4: 414; see p. 71 above). This notion *only* makes sense with respect to the influence practical reason might have on an imperfectly rational will that may also determine itself contrary to what practical reason prescribes in its imperatives. This idea is ruled out by the thoroughly Humean picture but explicitly endorsed by Kant: "Whoever wills the end, wills also (so far as reason decides his conduct) the means in his power which are indispensably necessary thereto" (GMS, 4: 417). The addition, "so far as reason decides his conduct" is significant here, because it signals that the influence that practical reason has on the will can be counteracted by sensible inclinations even in cases when the practical judgment that seeks to determine the will is not a moral judgment.¹⁸⁰ This addition suggests that there is a rational, normative influence of practical reason on the will that does *not* derive from moral considerations, and it avoids the implication that, with regards to non-moral ends, the question of what I ought to do collapses – leaving ignorance about empirical facts or causal relations aside – into the question of what I am actually motivated to do. If what I non-morally ought to do did wholly depend on my actual non-moral motives, I could not knowingly violate a HPJ, and then it would be wholly obscure on what basis I could represent a HPJ to myself as an 'ought'.¹⁸¹

¹⁷⁹ My actual motivation may be based on a false belief about the causal relations or facts. But here the relevant error is solely theoretical. See Korsgaard 1997 for a very forceful presentation of this point.

¹⁸⁰ This is recognized by Pollok 2007, who shows (pp. 62-3) that the addition, 'in so far as reason decides his conduct' is needed to give Kant's often merely descriptive claims about HPJ a normative upshot.

¹⁸¹ This is the decisive problem for exclusively Humean accounts of HPJ in Kant, such as Bojanowski's 2006. He tries to explain (p. 211) the 'ought-shape' of HPJ by appeal to two possibilities: (i) we may not know the means required for the realization of our non-moral ends; and (ii) we may be subject to motives that conflict with the realization of our non-moral ends. Now (i) only gives us the uninteresting case of a theoretical error that does not account for the knowing violation of a HPJ. With regards to (ii), Bojanowski states (a) that a HPJ loses its prescriptive force as soon as we give up the end that underlies it, and (b) that anyone who really wills the end automatically takes (what she represents as) the necessary means. He does not realize that the combination of (a)

This shows that Kant's claim that we always as a matter of natural necessity desire our happiness, qua uninterrupted contentment with our existence, cannot mean that whenever we act on the basis of non-moral motives we are motivated by our longing for happiness; for if that were what Kant meant the only possible error in the domain of non-moral willing would be theoretical (i.e., it would amount to ignorance about what makes us happy). Rather, I suggest that we should understand Kant as saying that any given point in our finite existence, the only thing we ever want (leaving moral interests aside) is pleasure; hence, as beings who are conscious of the temporal dimension of their existence we must always want to have pleasure through all of our future existence (i.e., we must long for an uninterrupted state of pleasure, which would be happiness in Kant's sense), or at least want the greatest pleasure for the longest possible time. But while this want for the greatest possible amount of pleasure affects us as a matter of natural necessity, it does not necessarily have motivational prevalence – not over our moral interests, but also (and more importantly for present purposes) not over desires for a kind of short-term pleasure whose satisfaction would come at the expense of our wish for long-term contentment.

Thus, when Kant says (KpV, 5: 23) that "the man who cares only for the enjoyment of life [asks] only how much and how great pleasure [objects] will give for the longest time", he is certainly aware that construed as a descriptive claim, this is false. What he puts forward here is, I suggest, the *normative* claim that reason *requires* that our non-moral choices be sensitive to the effect a choice has on one's overall satisfaction (i.e., on the 'best' overall distribution of pleasure over time). Practical reason puts forward this normative requirement because it has in view the demands of our sensible nature viewed as a whole, i.e., considered over an extended temporal stretch. It finds that being pleased with our own existence is the *only* thing (apart from moral ends) we care about, and that we always care about it equally at any given time; on this basis, it tries to channel our non-moral motives in a manner that corrects irrational biases toward the near future and that allows us to get what we want as long and as often as possible:

...as far as our nature as sensible beings is concerned, our happiness is the only thing of consequence, provided it is estimated as reason especially requires not by the transitory sensation, but by the influence that this has on our whole existence, and on our satisfaction therewith. (KpV, 5: 61)

For the human will is not determined by that alone which stimulates, that is, immediately affects the senses; we have the power to overcome the impressions on our faculty of sensuous desire, by calling up representations of what, in a more indirect manner, is useful or injurious. (CPR, A 802/B 830)

Here we must also go back to the aforementioned point that it is practical reason itself that deems an act that secures an agent's happiness good, and indeed represents such an act as possessing the only possible (conditional, instrumental) value that an act that is not directed at moral purposes could possibly possess. This verdict is irreducible to the influence of sensible desire in two respects. First, the ranking of an act directed at one's happiness in comparison to an act directed at moral purposes, as (only) conditionally good, could not come from our sensible nature. Secondly, the ranking of some specific desire 'd1' whose satisfaction would give us long-term contentment as being preferable to some desire 'd2' whose satisfaction would only give us short-

and (b) vitiates the attempt to explain the fact that agents represent HPJ to themselves as 'oughts' by appeal to (ii). For (b) implies that an agent who acts on a desire whose realization she knows to conflict with some end e does not really will e, and in conjunction with (a) this means that all it takes for an agent to negate the prescriptive force of a HPJ is to knowingly fail to take the means the HPJ deems required for e. For in this case she no longer wills e and hence the HPJ has no normative claim on her.

term pleasure could not come from our sensible nature either: the only way in which our sensible nature could favor (acting on) a desire for long-term pleasure is by allowing it to outweigh, in terms of strength or actual motivational force, (acting on) desires for short-term pleasure. But we often knowingly choose short-term pleasure at the expense of long-term satisfaction: what our faculty of sensibility inclines us to do on a given occasion is not fine-tuned to facts about what is most likely to make us happy in the long run, because it is determined by the influence of factors such as stress, exhaustion, or specters of the imagination. So there is no function from the *descriptive properties* of given empirical desires to the verdict that desires for something that would give us long-term contentment reflect our *real* ends, i.e., aim at some real rather than merely apparent value. Practical reason confers normative authority on our longing for long-term contentment and attempts to influence the will in cases where the strength of this longing wanes, by pronouncing HPJ that represent rules for the prudent, efficient pursuit of happiness. Hence, for Kant practical reason can motivate the will, even in the pursuit of non-moral value, in a manner that is irreducible to the contribution of desire.¹⁸²

In view of what I have argued thus far, it should be clear that Kant's attitude toward the Humean dictum that reason is the slave of the passions is deeply ambivalent, as far as our non-moral willing is concerned. I now want to summarize what I take to be the Humean and the anti-Humean elements in Kant's conception of non-moral willing.

I think that Kant's conception of non-moral willing is congenial to the Humean picture in three major respects. First, the concern with our happiness or continued satisfaction is not the result of autonomous (self-)legislation but rather imposed on practical reason 'from the outside', by our sensible nature. Much unlike the normative requirement that informs categorical imperatives, the normative requirement to choose acts that maximize our pleasure over time answers to the demands of our sensible nature. As we saw, practical reason here is at the service of our sensible nature considered as a temporally extended whole rather than at the service of its transitory configurations; but the fact that practical reason *is* at the service of our empirical nature is nonetheless a noteworthy concession to a Humean picture.¹⁸³ Moreover (this is the second Humean element in Kant's account), it is wholly contingent and rationally arbitrary what we non-morally ought to do in the deep sense that *whatever*, as a matter of contingent fact, maximizes our pleasure is what practical reason must recommend to us (provided that it be

¹⁸² How does this bear on Kant's somewhat quirky idea that securing one's long-term happiness can be an *indirect moral requirement* (since an unhappy finite being is more likely to violate the moral law)? In the Groundwork (GMS, 4: 399), Kant gives the following example: a rule of prudence says that P should renounce eating tasty foods, and P's general desire for happiness is insufficient to determine his will accordingly because he thinks that he can *count safely* on getting pleasure from the tasty food but not on getting pleasure from being healthy in the future. Here, Kant suggests, the influence of the moral law (the indirect duty to do what actually makes one happy) can determine P to renounce the tasty food in favor of his future health. I am not usually quick to point out that Kant is confused, but here I cannot help thinking that he is. If P honestly but incorrectly believes that enjoying the food is more like to make him happy than being healthy, then the indirect duty to do what makes him happy cannot move him to renounce the food, because the moral law here has to go via P's representation of what it is that would make him happy. The moral law cannot correct errors in means-end reasoning. Conversely, if P became convinced that future health is more likely to make him happy than enjoying tasty food, then the non-moral norm of prudence which says that he ought to renounce the food might move him to renounce the food. Kant's point might be that if the rule of prudence *in this case* failed to move P, the moral law might make a separate motivational contribution.

¹⁸³ See KpV, 5: 61-2 for a very clear statement to the effect that all judgments of practical reason, except for moral judgments (i.e., all HPJ), are the service of our sensible nature.

morally permissible). Finally (third), and precisely because of this dependency of non-moral practical reason on our given psychological constitution, practical reason depends for answering the question of what our non-moral ends ought to be on the deliverances of theoretical reasoning. Theory discovers, within limits, causal facts concerning what would actually maximize our overall contentment with our existence.

On the other hand, Kant's conception of non-moral willing also contains a couple of strikingly anti-Humean elements. Although our longing for happiness as well as our conception of what makes us happy are imposed on us by our dependent empirical nature, for Kant the verdict that an act which helps a dependent rational to realize its inescapable longing for happiness is (pending its moral permissibility) worth pursuing stems from practical reason rather than from our 'lower' faculty of desire. A genuinely *practical* judgment, which may have an impact on the will, requires that practical reason takes up the material provided by theoretical reason, supplies its own *sui generis* concept of conditional value, and represents taking the route toward happiness that was discovered by theoretical reasons as instrumentally *good*.¹⁸⁴ Moreover, from the fact that we are caused, by our empirical nature, to want happiness it does not follow that we are forced or caused to set ourselves any specific end: setting ourselves ends is always an act of freedom or spontaneity. This is so for two reasons. First, we are ('negatively') free to subordinate our longing for our own happiness to our concern with moral values or vice versa. Second, the end of happiness is a merely formal end that lacks specific content; to set ourselves specific non-moral ends, we must determine (through theoretical reasoning) which of the various elements in the manifold of our given desires is worth acting upon, i.e., relates to our long-term contentment (or happiness) rather than our mere short-term pleasure. Since theoretical reason can discover some facts about the routes that lead toward happiness, practical reason deems following those routes conditionally good and rational, independently of and in possible conflict with what an agent may happen to desire at any given point in time. So for Kant even non-moral practical reason is not the slave of a *momentarily ruling* passion: a finite agent's faculty of will or practical reason does not, even apart from the concern with moral ends, collapse into desires that affect her fragmented empirical self across time. This is an important anti-Humean aspect of Kant's conception of non-moral practical rationality.¹⁸⁵

Even the claim that Kant's conception of non-moral willing involves a significant anti-Humean strand stands in need of an important qualification. For Kant it is merely contingent that

¹⁸⁴ In the First Introduction to the Critique of Judgment, Kant claims that those 'practical statements' which specify means to our ends are distinguished from theoretical statements only according to their manner of representation, not according to their content (EEUK, 20: 196): they qualify as 'practical' only in virtue of their formula, not in virtue of their content (197). I think he overstates his case here: the content of a HPJ is not exhausted by a theoretical cognition, because the HPJ does not just say that one must do x in order to produce y, but also that doing x is (instrumentally, conditionally) *good*; concepts of value cannot be supplied by theoretical reason. It is only because of the inclusion of this evaluative *content* that the practical 'ought' formula, which as such is designed to subject the will to rational influence, has any application; this is the point of Kant's insistence that a hypothetical imperative represents an action *as good* (GMS, 4: 414). So, pace Cramer 1972, p. 196, I think that a HPJ can manifest itself as a practical prescription by virtue of its formula only if it contains a characteristically practical concept, that of (conditional) goodness, in its content.

¹⁸⁵ Note that (unlike Hampton 1998 and Korsgaard 1997) my argument for an anti-Humean element in Kant's conception of HPJ does not draw on the idea that hypothetical imperatives presuppose categorical imperatives. This idea is precluded by the fact that Kant says that the existence of hypothetical norms does not suffice to verify the existence of categorical norms (GMS, 4: 419). I will return to this point in chapters V and VI.

practical reason makes a contribution to our non-moral willing that is irreducible to the contribution of empirical desire. It is perfectly conceivable that we may be so constituted that our natural instincts infallibly incline us to do those things that make us happy.¹⁸⁶ If our natural instincts reliably inclined us to act in ways that lead to lasting contentment, non-moral practical reason would collapse into empirical desire.

This concludes my attempt to reconcile the various Humean and anti-Humean elements in Kant's conception of the role of practical reason in non-moral motivation. I now want to discuss how my interpretation bears on Kant's idea that HPJ are *analytic*, and what it implies for the normativity of HPJ. The first thing to note here is that what is analytic is not the *descriptive* claim that if a rational agent fully wills an end he thereby also wills the means. Rather, the analyticity pertains to a prescription or command: "the imperative which ordains the willing of the means to him who wills the end is...analytical" (GMS, 4: 419).¹⁸⁷ So, Kant's idea is that the very concept of willing an end contains the requirement to take means that one knows to be conducive to realizing this end (cf. GMS, 4: 417). But the second thing to notice here is that Kant holds that we can satisfy or, as he puts it, 'get rid of' ("los sein können") this prescription *without* taking the means we know to be necessary to our ends: namely, by abandoning the end (cf. GMS, 4: 420).¹⁸⁸ And this makes it look as if Kant is thinking of instrumental rationality as consisting of a so-called 'wide-scope requirement' that the agent can satisfy by choosing to either give up his end or to take the means. On this interpretation of the 'normative logic' of HPJ, which is suggested by some of the most perceptive commentators,

what the Hypothetical Imperative condemns is not the changing of one's aims; it is the irrationality of continuing to profess, work toward, and hope for certain ends even though one is unwilling to take some essential means to realize them.¹⁸⁹

If this interpretation were correct, the specific kind of irrationality associated with violating a HPJ would, on Kant's view, be that which is exhibited by someone who half-heartedly 'pursues' a certain aim, that is, continues to adopt the aim in some sense (it is actually a bit hard to see what this sense is) but fails to take the means she knows to be required for realizing it.

I think that if we take into account the broader context of Kant's theory of practical normativity, a rather different picture emerges. As should be clear from my argument thus far, I think that we can only understand Kant's conception of different kinds of practical norms (CPJ and HPJ) by taking into account his conception of different kinds of values. Each type of practical judgment relates to a distinctive kind of value from which it derives its reason-giving, normative force. Since for Kant the sphere of instrumental rationality is delimited by the one overarching end of happiness, all HPJ must derive their reason-giving force (ultimately, and maybe only mediately, as explained above) from the value of happiness. Since the value of happiness is only conditional, and subordinate to the unconditional moral value of dignity, it

¹⁸⁶ See GMS, 4: 395; KpV, 5: 61-2. In this case, HPJ would not be imperatives. I return to this point in chapter VI.

¹⁸⁷ This is a consequence of the fact (cf. footnote 180 above) that valid HPJ motivate the will not 'automatically' but only under the added condition that (practical) reason decides our conduct (i.e., that we are moved by *practical norms*). That this added condition is essential to the analyticity of HPJ is emphasized by Cramer 1972, p. 164, Korsgaard 1997, p. 238, and Pollok 2007, p. 69.

¹⁸⁸ See Pollok 2007, p. 61.

¹⁸⁹ Hill 1973, p. 435. This diagnosis also fits Korsgaard's discussion of 'Timid Prudence' (1997, p. 236 ff.). The wide-scope interpretation is most explicitly endorsed by Pollok 2007, p. 65. For discussions of the wide-scope/narrow-scope distinction see Broome 2007 and Kolodny 2005.

follows that the existence of a reason to perform a certain act that is likely to promote the agent's happiness is conditional upon whether the act does not conflict with the moral value of dignity.

Approaching Kant's conception of practical (including instrumental) normativity from what one might call a 'value-theoretic' perspective provides a neat explanation for why there can be no conflicting requirements of practical reason. Suppose I will to shoot a man (because I know that doing so is likely to make me happy). Here it is abundantly clear that Kant holds that the *only* rational option for me is to drop the end (rather than taking adequate means, such as buying a gun). My interpretation accommodates this point by emphasizing that there can only be a valid HPJ if (i) there is a course of action that is likely to secure the agent's happiness and if (ii) this course of action does not conflict with the absolute value of human dignity. Proponents of the wide-scope interpretation may offer their own account of why abandoning an immoral end is the only rational option: for instance, they can hold that the hypothetical imperative provides me with an alternative between two options and the categorical imperative eliminates one of these options.¹⁹⁰ But this suggests that *there is a valid HPJ* even when the agent acts on the basis of a morally bad maxim, and this ignores the fact that for Kant HPJ as such relate to a distinctive kind of goodness whose existence requires the moral permissibility of the agent's maxim; and hence the validity of a HPJ depends on whether the act that the HPJ represents as good is morally permissible. Moreover, proponents of the wide-scope interpretation face a difficulty in trying to make sense of cases where *a HPJ itself* recommends one alternative over the other. Suppose, for instance, that I will an end the realization of which is likely to make me happy and whose pursuit is morally permissible. Practical reason here is not indifferent between the options of pursuing the means and of dropping the end, but informs me, *via a valid HPJ*, that taking the means is good and that I ought to take the means, since this is the more likely way to realize my own happiness (alternatively, since is the more likely way to realize the only non-moral, formal end that my sensible nature – considered as a whole, in the way specified above – contains).

We can see the relevance of these issues by asking how Kant's conception of instrumental rationality is supposed to avoid the problem of 'bootstrapping' that afflicts any view according to which the mere fact that an agent has a 'committed herself' to an end gives her a reason to keep pursuing the end.¹⁹¹ Even if we leave out of account cases where pursuing the end is morally impermissible, it is unclear how, from my first-person deliberative perspective, I can see a point in pursuing an end if the only thing I can say in favor of doing so is that I have in fact adopted the end.¹⁹² Now, proponents of the wide-scope interpretation tend to claim that it is a virtue of their account that it helps to avoid this problem, because on their view the mere fact that an agent has adopted an end does not entail that she ought to take (necessary) means toward this end.¹⁹³ But it seems to me that the wide-scope interpretation faces a dilemma here. Either the wide-scope norm that one ought to either give up the end or take the means implies that there is a reason for giving up the end and also a reason for taking the means (since both options are ways of satisfying the wide-scope requirement), but (this is the first horn) in this case there is

¹⁹⁰ This is Hill's solution (1973, p. 436).

¹⁹¹ For lucid discussions of bootstrapping, in relation to instrumental reason, see Bratman 1987 and Kolodny 2005.

¹⁹² See, for instance, Korsgaard's idea that "your willing the end gives it a normative status for you...your willing the end in a sense makes it good" (1997, p. 246). She rightly emphasizes the importance of the first-person perspective to the phenomenon of valuing and recognizing reasons; but it is not clear how one can regard, from this perspective, the *mere* fact that one wills an end as a reason for pursuing the end or as a ground of goodness.

¹⁹³ See, for instance, Hill 1973, pp. 435-6, and Pollok 2007, p. 65.

considerable bootstrapping after all (and, moreover, it is hard to see how an agent can take seriously her choice between abandoning an end or pursuing it, if both options are rationally on par). Alternatively, the proponent of the wide-scope account might say that whether one has reason to take the means or drop the end is independent of instrumental reason itself, which merely tells one to do one or the other; but (this is the second horn) in this case the principle of instrumental reason seems normatively idle – a deliberating agent can, and should, bypass the wide-scope norm and focus directly on whatever considerations are sources of reasons. The second horn is avoided by Hill's proposal that the central phenomenon which the principle of instrumental reason seeks to address is that of half-heartedly 'pursuing' an end without real commitment: for this proposal preserves a genuine normative point for the instrumental principle. But if preventing such half-heartedness is the source of the normativity of instrumental reason, the problem of bootstrapping keeps looming, for in this case I do as I ought to as long as I am wholehearted in my pursuits, i.e., as long as I either efficiently pursue the end or wholly abandon it whenever I don't feel like taking the means – *completely regardless of any considerations concerning the end*. Apart from the philosophical worries this proposal raises, I see no textual evidence that supports the idea that Kant's conception of instrumental reason centers around the phenomenon of half-heartedness. Rather, it seems that Kant's account is centrally informed by the idea that for dependent beings like us, our empirically given interest in our own happiness is the one and only source of instrumental reasons (i.e., of reasons for actions that are good as means, for something other than themselves, "wozu anders"; GMS, 4: 414).

What, then, does my interpretation make of the analyticity of hypothetical imperatives? There has been a debate about whether Kant acknowledges a general form of 'the' Hypothetical Imperative over and above specific HPJ.¹⁹⁴ I accept that Kant posits a general form of the Hypothetical Imperative, namely, the analytic principle (which we can establish by analysis of the concept of 'willing an end') that if one wills an end *e*, and if one knows that doing *m* is necessary for realizing *e*, then one ought to do *m* or give up *e*.¹⁹⁵ But I do not accept that for Kant this analytic principle is, as such, a source of instrumental practical reasons. That this principle lacks the normative significance that commentators tend to assign to it is suggested by the fact that it plays absolutely no rule (is not even mentioned) in Kant's discussion of non-moral practical reasons in any of the three Critiques. But it is already in the Groundwork that Kant implies that the general analytic form of hypothetical imperatives has no normative purchase on a concrete volition and fails to specify what an agent ought to will, unless the nature of the end is specified: he states that when I conceive of a hypothetical imperative "in general", I do not know what it contains unless "the condition" is 'given' to me. To conceive of a hypothetical imperative in general is to conceive of the general form of the Hypothetical Imperative. This form cannot ground a genuine prescription until we add a concrete end ('the condition').¹⁹⁶ Now, if we focus our attention on finite rational agents, we can see that although the abstract general form of the Hypothetical Imperative can be considered analytic, *there are no concrete analytic HPJ*. This follows from the combination of two ideas: first, the only end that can be a source of non-moral reasons and hence of valid HPJ addressed to human agents is (their own) happiness; second,

¹⁹⁴ The view that there is no general hypothetical imperative form is held by Ludwig 1999, p. 107. The view that there is such a general form is defended by Hill 1973, Schönecker/Wood 2002 and Pollok 2007.

¹⁹⁵ This is very similar to Schönecker/Wood's 2002, p. 111 formulation, except that I have added 'or give up *e*' in view of Kant's remarks at GMS, 4: 420.

¹⁹⁶ See GMS, 4: 420.

human agents do not know necessary means for realizing this end. Since the specific HPJ addressed to human agents fail to represent necessary means, they fail to fully instantiate the general, analytic form of the Hypothetical Imperative.¹⁹⁷

This draws attention to an important and controversial consequence of my interpretation: On my account of HPJ, neither the standards of non-moral willing (as discovered by theoretical reason) nor the practical norms that represent those standards to the will involve any genuine necessities.¹⁹⁸ There can be no *strictly necessary* rules of practical rationality in the non-moral domain, for the reasons provided earlier: the value represented by HPJ is merely conditional, and the rules (counsels of prudence) for realizing this value lack universality and certainty. Now, Kant does, at least in the Groundwork, suggest that imperatives of skill can express genuine necessities (and, correspondingly, that they can be considered analytic). His point here is surely that theoretical reason can establish causally necessary means to some *possible* concrete end. But he fails to point out here that imperatives of skill *practically presuppose* counsels of prudence and hence inherit those features that prevent these counsels from expressing a practical necessity. An imperative of skill that represents (say) covering the structure as necessary for building a hut can have a purchase on my will and strike me as a prescription of practical reason ('It would be good/I ought to cover the structure') only if I believe that building the hut is going to make me happy. Since there is only a certain likelihood in the notion that building the hut is going to make me happy, the idea that I ought to cover the structure cannot express a genuine necessity; moreover, even if I knew that building the hut would make me happy, the relevant HPJ would still fall short of expressing a strict practical necessity because (first) it would not apply to all agents who strive for happiness and (second) because the merely conditional value of happiness cannot ground unconditional practical necessities. Kant acknowledges these facts in the Second Critique, where he says of hypothetical imperatives *simpliciter*, conceived as 'counsels', that they lack necessity (KpV, 5: 20; 63).¹⁹⁹

From the fact that there are no genuine necessities of non-moral practical reason, it does not follow that HPJ are pointless, or that any morally permissible act performed for the sake of non-moral ends is as good or rational as any other. The advice provided by HPJ may be inconclusive and defeasible, but it is genuine advice that it would be irrational to dismiss.²⁰⁰

It can be objected that my interpretation misconstrues the point of imperatives of skill by making them practically and normatively dependent on counsels of prudence. On an alternative

¹⁹⁷ See the following counterfactual at GMS, 4: 419: "This imperative of prudence would however be an analytical proposition if we assume that the means to happiness could be certainly assigned."

¹⁹⁸ For the insistence that a HPJ must tell us that certain acts are necessary, see Hill 1973, p. 435.

¹⁹⁹ Compare KpV, 5: 25-6, where Kant says that a principle such as, 'someone who wants to eat bread has to devise a mill' has genuine universality (never mind his oversimplification here), but only as a theoretical principle; as a practical principle (and hence as a HPJ) it must lack such universality (and necessity) because the extent to which agents have the desire for bread (and, we can add, the extent to which satisfying this desire is conducive to their long-term happiness), is empirically contingent.

²⁰⁰ Here I disagree with Herman 2001, pp. 140-41 who claims that *only* moral willing makes room for weakness of will. She seems to infer from the fact that rules of prudence provide only defeasible advice that violating such rules is not irrational. But Kant himself deems the notion that it is advisable to save money in one's youth so that one doesn't have to struggle in one's old age a "correct and important prescription of the will" (KpV, 5: 20). Herman's interpretation seems to be based on what Kant says at KpV, 5: 23; but Kant's point here is only that all objects of non-moral deliberation are to be measured by the one standard of how much pleasure they would give us. It does not follow that if I make a choice contrary to established standards of happiness I cannot be considered irrational. See KpV, 5: 62 for Kant's emphasis that (mere) practical counsels can be *vernünftig*.

account of imperatives of skill, the point of these imperatives is *merely* to guide purposive behavior in the production of objects of a certain kind. As such, they would express an 'ought' that is, as Hannah Ginsborg puts it, "thinner than the 'ought' of rational necessity"²⁰¹ in the sense that the 'ought' they express tolerates the agent's sense of the *practical irrationality* of the action. For instance, if I want to produce an omelette, "I take my actions to be normatively governed by the concept of an omelette, and say to myself things like 'now I ought to beat the eggs', even if I recognize that the project as a whole is irrational" (because I engage in it "out of a neurotic compulsion").²⁰² This objection to my construal of imperatives of skill correctly points out that there is a *standard* that says that beating the eggs is the *appropriate step to take* given the fact that the intended end product is an omelette. Ordinary language certainly does contain a sense of 'ought' that relates to nothing more than the existence of such a standard; and Kant's discussion of imperatives of skill sometimes concerns precisely this sense of 'ought'.²⁰³ But I want to insist that in the context of an analysis of *prescriptions of practical reason*, Kant cannot be concerned with this 'thin' sense of 'ought'. This is so for two reasons. First, notice that the sense of 'ought' that merely relates to the existence of some technical standard applies even to the workings of a mere mechanism that has no faculty of practical reason (imagine a sophisticated robot that is programmed to make omelettes). More importantly, if Kant did think that the 'ought' deriving from technical standards could all by itself ground a HPJ, he would be committed to saying that practical reason issues conflicting prescriptions. This is obvious when we look at a case where there is a technical standard governing an *immoral* action (e.g., the assassin's mixing the poison 'as she ought to'), but it is equally obvious in the case in which I make an omelette out of a neurotic compulsion: in this case there is a valid HPJ stating that I ought not (given my end of happiness) to beat the eggs. A HPJ, qua prescription of practical reason, must relate to the agent's will, that is, to what it would be good for her to do. Hence, whether I *ought* – in the sense relevant to a HPJ qua prescription of practical reason – to take necessary means toward a concrete end depends on whether realizing this end is compatible with the absolute value of moral goodness and likely to realize the conditional value of my own happiness.²⁰⁴

²⁰¹ Ginsborg 1991, p. 18.

²⁰² Ginsborg 1991, p. 18.

²⁰³ Kant's claim that in the case of imperatives of skill "there is no question whether the end is rational and good, but only what one must do in order to attain it" exactly fits Ginsborg's account. I think that what Kant says here is correct when we view rules of skill 'out of context', namely, in abstraction from their relation to our willing and our deliberative pursuit of happiness.

²⁰⁴ In § 10 of the Critique of Judgment, Kant talks about the purposive behavior that results when an object is produced in accordance with a concept (qua rule of production). His exposition contains a very puzzling remark: "The representation of the effect is here the determining ground of its cause and takes the lead of it." A further interpretative puzzle about § 10 is raised by the fact that Kant mentions the will qua capacity to desire what is pleasant, before he gets to the systematic point of his discussion ('purposiveness without a purpose'). I think we can use the account of imperatives of skill that I have been proposing to make sense of these remarks. When one is moved by the prospect of pleasure, the representation of a pleasure that would result from some action is the determining ground of and precedes one's act. One brings into existence an object because one thinks that this will cause pleasure. The action is the cause, and the pleasure is the effect. Now, the representation of the effect (i.e., of the pleasure) can be said to be the determining ground of the action, because it is for the sake of that expected effect that one performs the action in the first place. This is surely the point of Kant's remark that in producing an artifact, one is not concerned with the mere cognition of an object, but with the object itself, 'its existence': there is a connection between the interest one takes in the existence of an object and the anticipation of pleasure (see §3-5 of the Critique of Judgment). There are two different kinds of effects of a production: first,

We are now in a position to specify how Kant conceives of the meaning of HPJ. The content of HPJ relates to a conditional value and does not express a genuine necessity. Neither of these features prevents HPJ from representing practical reasons and norms. A norm relating to a conditional value applies only under certain conditions, but if those conditions obtain it is a genuinely valid norm. And even if the norms expressed by HPJ cannot specify with any certainty what we have to do in order to realize the value of our own happiness, they can specify which actions are more likely to bring about the valuable effect; if action a is more likely to realize a valuable effect than action b, there is a norm of practical reason which advises one to perform a rather than b. Thus, both CPJ and HPJ represent norms of practical rationality. As we saw, the second feature that accounts for the meaning of a practical judgment is its *formula*; and here we can see that the formulae in which HPJ and CPJ represent their characteristic norms are identical. HPJ are addressed to agents who, as a matter of natural necessity, desire the one thing of non-moral value, their own happiness, but who are not necessarily motivated to act in ways that increase the likelihood of actually realizing this value, i.e., whose actual, empirically given desires are not reliably attuned to what in fact would give them the greatest net sum of pleasure. Hence, there is a sense in which "[practical] reason will not give way to any ground which is empirically given" even when "what is willed [is] an object of mere sensibility (the pleasant)".²⁰⁵

3: Kant's incompatibilism about practical normativity: two distinct worries

Now that we have a sense of how Kant conceives of the meaning of practical ought judgments, I want to give a preliminary answer to the question of why Kant thinks that if all causality were

the existence of the artifact, which is an effect of the concept-guided productive activity; second, the pleasure that the producer gets (however mediately) from bringing the artifact into existence. Production governed by the concept of the artifact is the cause of the existence of the artifact; and the artifact is produced in the first place because the producer takes an interest in the existence of the artifact, which is to say that he expects pleasure from the production of the artifact. The representation of this pleasure is the determining ground of the concept-guided activity of producing the artifact, because it is for the sake of this pleasure that the artifact is produced. This interpretation makes good sense of the puzzling remarks in § 10, and it confirms the idea that the practical relevance of imperatives of skill is essentially tied to the pursuit of happiness (pleasure).

²⁰⁵ One might object to my account of the normativity of practical judgments that it violates Kant's demand for the unity of reason, because on my interpretation the source of normativity of HPJ is a given empirical desire (for happiness) whereas the source of normativity of CPJ is practical reason itself, independently of any given empirical desire. Pollok claims that given Kant's commitment to the unity of reason, it would be 'surprising if Kant did posit different sources of practical normativity for different forms of imperatives' (2007, p. 72). I want to make a couple of points in response. First, it is not clear to me that Kant's commitment to the unity of reason does not tolerate a variety of sources of normativity – I think that this commitment is driven first and foremost by the demand that the various claims of reason (e.g., practical reason's claim that we are free and theoretical reason's claim that we are causally determined) do not stand in conflict (or even, as he argues in the Second Critique, that the idea of freedom supplements and assists the theoretical use of reason). Second, and relatedly, my account does preserve the unity of practical reason in this sense because it avoids a clash between HPJ and CPJ, or between the two distinctive values associated with HPJ and CPJ. Third, it seems to me that Kant's commitment to the unity of reason relates only and strictly to *pure* reason – that is, to the unity of pure practical and pure theoretical reason. As Kant never tires of emphasizing, practical reason in its non-moral, instrumental employment is not pure but conditioned by our empirical faculty of desire (KpV, 5: 21; 23-4), and from the standpoint of instrumental rationality (concerned with our own happiness) we would be much better off if we were governed by instinct rather than by reason (GMS, 4: 395). Finally, as I argued above, I do not think that Pollok's attempt to show that HPJ and CPJ have a common root succeeds, because his attempt to construe HPJ as a wide-scope requirement of practical reason independently of desires implies a kind of bootstrapping that is alien to Kant and that fails to accord with this conception of non-moral value.

mere nature, practical ought judgments would be *meaningless*. I want to suggest that Kant's incompatibilism about practical normativity draws on two separate considerations: if all causality were mere nature, practical judgments would neither have the *content* nor the *formula* which jointly account for their distinctive meaning.

In chapter V, I will expound and review the kinds of considerations that lead Kant to suggest that practical judgments would not have the kind of *content or validity* we intend them to have if all causality were mere nature. To anticipate, Kant's worry is that if all causality were mere nature, our practical judgments would be prompted by states and processes that cannot imbue them with the kind of normative, rational authority they purport to express. "Sensuous impulses may impel me to will", but "they can never give rise to the 'ought', but only to a willing which, while very far from being necessary, is always conditioned" (CPR, A 548/B 576). The worry concerns a misfit between the normative content of practical judgments and the supposition that such judgments are the result of the kinds of states and processes that exhaust our constitution under the assumption that we are wholly parts of the contingent natural order.

In chapter VI, I will expound and review the kinds of considerations that lead Kant to suggest that if all causality were mere nature, the 'ought' formula of practical judgments would lose its point. To anticipate: the notion that all causality is mere nature implies that all our actions result exclusively from deterministic processes. If so, any action that is possible under natural laws is also necessary under those laws (i.e., the laws of nature do not preclude the occurrence of such an action only if they necessitate its occurrence). This conflicts with the 'ought' formula, which requires that an agent can comply with a norm without being necessitated to do so.

My idea that these are two separable worries about the idea that all causality is mere nature has several implications. First, this idea is connected to Kant's claim that the will of finitely rational beings divides into two conceptually separable sub-faculties, one that is responsible for law-giving (will qua practical reason) and one that is responsible for volitional structure (Willkür); see MS, 6: 226. The freedom of practical reason that we take ourselves to exhibit when we legislate practical norms requires that we can make practical judgments without depending on the influence of sensible conditions – this is part of what Kant means when he says that *we cannot act but under the idea of freedom*²⁰⁶ – but it does *not* require the capacity to 'judge otherwise'. The commitment to 'alternative possibilities' is characteristic of our conception of a free Willkür, i.e, a capacity for of agency that requires a 'double' capacity to comply with or to violate practical norms in the formation and execution of volitional structures.²⁰⁷

A second point that is worth noticing here is that Kant's second worry, concerning the implications that the assumption that all causality is mere nature has for the *formula* of practical judgments, does not pertain to the possibility of practical normativity as such. That is: the domain of possible agents who are governed by normative laws is broader than the domain of agents who are governed by 'oughts'. It is conceivable that there might be a perfectly rational

²⁰⁶ GMS, 4: 448: "Now I affirm that we must attribute to every rational being which has a will that it has also the idea of freedom and acts entirely under this idea. (...) Now we cannot possibly conceive a reason consciously receiving a bias from any other quarter with respect to its judgements, for then the subject would ascribe the determination of its judgement not to its own reason, but to an impulse. It must regard itself as the author of its principles independent of foreign influences. Consequently as practical reason or as the will of a rational being it must regard itself as free, that is to say, the will of such a being cannot be a will of its own except under the idea of freedom." See my chapter III, section 3 for further discussion of Kant's concepts of 'actions'.

²⁰⁷ See, again, chapter III, section 3 for an elaboration and defense of these points.

('holy') will which *necessarily* does what practical reason requires. As we saw, it would be pointless to address a perfectly rational being in terms of 'oughts'. One might object that this reduces the significance of Kant's second worry, since it only relates to a condition on the manner in which the normative content of practical judgments is represented to beings like us.²⁰⁸ I think Kant would respond by insisting that it is a deep and essential fact about human agents that they are only imperfectly rational and (hence) that the only possible way in which normative principles can influence our will is via an 'ought'. Our entire practical discourse is, without exception, mediated through the *ought* formula, because our entire practical discourse is characterized by the idea that correct normative principles can but also may fail to determine us to act. Furthermore, it is important to be clear about the fact that normative principles play an altogether different role for us and for a holy will. The difference between us and a perfectly rational will is not simply that such a will is immune against any temptation to flout a correct practical norm. The conception of such a will is more radical in the sense that it even excludes the possibility of there being different options about what to do, and hence it excludes the predicament of practical deliberation and choice. Such a being is never even faced with more than option about what to do, because every choice of such a being coincides, of necessity, with what the moral law demands (GMS, 4: 413-4). By contrast, for beings like us, the relevance of practical principles depends on their impact on our deliberation and choice. Hence, the mere notion of a practical principle as such, irrespective of its formula, is very thin and it leaves out of account features that are indispensable to our self-conception as (sensibly affected) agents.

Third, it must be noticed that the distinction between two conceptually distinct worries about idea that all causality is mere nature is never made explicit by Kant. My distinction is between a 'naturalistic' worry about the idea that our practical judgments are prompted by natural states and processes and a 'deterministic' worry about the notion that our actions are necessitated by antecedent states and processes. Kant does not explicitly distinguish between these two worries because for him it is constitutive of the very possibility of a system of nature that this system is deterministic (i.e., that natural events occurs as a matter of material, law-like necessity). Hence, for Kant determinism and naturalism are two sides of one coin.

I want to end this chapter by addressing two questions that one might raise about Kant's picture. First, why should it follow from the idea that 'all causality is mere nature' that all practical judgments are *meaningless rather than simply false*?²⁰⁹ That is, why should we think that the assumption that all causality is mere nature implies that normative judgments *fail to express* that we ought to do certain things, and not simply that it is *never the case* that we ought to do one thing rather than another? Here two points are relevant. If we focus on the *origin* of practical judgments, I think that Kant's point is that a *false* practical judgment has a kind of negative normative significance that it cannot inherit from natural causes. The notion of a false practical judgment implies the notion of an *error* of practical reason: i.e., it implies that practical reason endorses a motive as worth acting upon that is not in fact worth acting upon. This, in turn, presupposes that there is room for a (inappropriate) contribution of reason that is not reducible to the influence of those states that constitute the domain of 'natural causality'. Here it might be helpful to consider, by analogy, what it takes for an action to have the kind of negative

²⁰⁸ This seems to be an upshot of Lavin's 2004 discussion. He states that it is "plainly and trivially true that an agent is subject to an imperative only if it is possible for the agent to violate it" (p. 441); he seems to suggest that this 'trivial' point is uninteresting because it fails to tell us something about practical normativity *as such*.

²⁰⁹ Pereboom 2005 claims that Kant thinks that determinism would imply that all practical judgments are false.

significance that invites reactions such as blame and reproach. This presupposes that the action is not solely due to the influence of desires, but that such influence was *granted to* these natural motive causes by the very same faculty of will that allows the agent to resist the influence of desire and to perform an act that has *positive* significance (i.e., that merits praise). If the action were to be attributed solely to contingent desires, without any contribution of practical reason, the action could not be represented as possessing the kind of negative significance that renders it an appropriate object of blame. I think that Kant would say that the same point applies, *mutatis mutandis*, to the possibility of erroneous practical judgment.²¹⁰ There is a second point here that emerges when we focus on the *object* rather than the origin of practical judgment: when Kant says that "when we have the course of nature alone in view, 'ought' has no meaning whatsoever", part of what he means is surely that actions, considered solely as the result of deterministic processes, are not the types of things about which we can make ought-judgments. We might read him here as indicating that if one tries to subject natural events to normative prescriptions, one is guilty of a kind of category mistake which renders one's judgments point – or meaningless.

Second, one might wonder why we should think that Kant's worries about the implications of the idea that all causality is mere nature relate to the conditions on the meaning of practical judgments rather than to the conditions on *rational agency*. In response, we can first note that for Kant practical reasoning itself is a form of rational agency (cf. *GMS*, 4: 448; see my chapter III, section 3) aimed at the formation of practical judgments. Second, when we focus on 'action' in a more narrow sense as including both deliberative choices and the performance of 'outer' acts, it seems that such acts may have a rational, normative dimension *just in case* they can be influenced by and assessed in terms of concepts such as 'good, bad, rational, irrational, legitimate, required', concepts that form the content of practical ought judgments. If there are no meaningful ought judgments, our behavior lacks the kind of rational, normative dimension that it owes to its being governed by meaningful judgments of this type. Hence, we can say that the notions of practical judgment and rational agency (in the specified, narrow sense) are *interdependent*: the object of practical judgment is an intentional (inner or outer) action, and hence one cannot get a grip on the content of practical judgment without appealing to the notion of intentional action; on the other hand, since intentional action owes its rational dimension to the possibility of its being influenced by and assessed in terms of the norms represented by practical judgments, we cannot get a grip on the rational dimension of intentional agency without appealing to the notion of practical judgment. Whether one adopts a compatibilist or an incompatibilist account of the conditions of practical normative judgment will determine whether one adopts a compatibilist or an incompatibilist account of the conditions of rational agency.²¹¹

²¹⁰ For Kant's claim that the appeal to sensible conditions or desires is insufficient to ground the concept of evil, see *RGV*, 6: 21, 34-5; for the – I think – analogous claim that such an appeal cannot ground the concept of erroneous judgment, see the *CPR*, A 394-5/B 350. I will return to this analogy in chapter IX.

²¹¹ One might raise the following worry about Kant's position: when he elucidates why practical normativity requires an incompatibilist kind freedom, he appeals to the idea that rational activity requires incompatibilist freedom; but if he in turn motivates incompatibilism about rational activity by appeal to the idea that rational activity requires an incompatibilist notion of practical normativity, he is caught in a vicious circle. In response, I want to say that since the notions of practical normativity and of rational activity imply one another (i.e., we cannot get a grip about what it is to engage in rational activity without invoking the notion of practical normativity, and vice versa), we can, without circularity, motivate incompatibilism about practical normativity by appeal to incompatibilism about rational activity. We need not suppose that these are two conceptually distinct kinds of incompatibilism that require independent motivation. (The same goes, *mutatis mutandis*, for a

My last clarification here concerns the fact that of the two incompatibilist worries that I ascribe to Kant, the first one relates in a straightforward sense only to CPJ. That is, the worry that if practical judgments were formed under the influence of contingent empirical desires they could not come to express a strict practical necessity is inapplicable to HPJ, for such judgments, as we saw, do not express practical necessities. However, an analogous worry might apply even to HPJ. HPJ are, as we saw, are in the service of our sensible nature and hence represent desire-dependent reasons. Nonetheless, the normative logic of those judgments does seem to presuppose that there is a meaningful distinction to be drawn between desires that have and desires that lack normative authority, or between what an agent's strongest desire is at any given moment (i.e., what she actually does) and what an agent has reason to do. So even HPJ purport to have a kind of normative authority over given desires; and Kant's first worry (which he explicitly only ever raises with regard to CPJ), as applied to HPJ, might be that if all practical reasoning and judgments stood exclusively under the influence of desires, they could not be said to have such authority. I shall explore this idea in the following chapter.

compatibilist conception of practical normativity and rational agency.)

Chapter V: Naturalism and Normative Practical Authority

In the preceding chapter, I argued that Kant's idea that practical 'ought' judgments would be meaningless if 'all causality were mere nature' divides into two separate arguments. In this chapter I will try to reconstruct and assess the first of these arguments, which is centered on the idea that a naturalistic view of human reason implies that our practical judgments cannot have the kind of *content, validity or authority* that we tend to ascribe to them.

1: The intuitive foundations of Kant's anti-naturalism about practical normativity

The argument I am attributing to Kant is contained, in extremely compressed form, in his claim that "natural grounds [such as] sensuous impulses may impel me to will [but] can never give rise to [hervorbringen] the 'ought', but only to a willing which, while very far from being necessary, is always conditioned" (CPR, A 548/B 576). Two things are worth emphasizing here. First, Kant's point in this passage is straightforwardly causal: both the notion of *production* and the appeal to *natural grounds* make clear that he is concerned with the *efficient causes* that are responsible for practical reasoning and its outcome (i.e., representations such as practical concepts and judgments).²¹² A naturalist believes that *all* relevant causes of events or states must be such that their existence can (at least indirectly) be verified on the basis of sense experience and that they have a place in the explanations of natural science (cf. p. iii of my Introduction). We should conceive 'natural science' here broadly, in a non-reductive sense, as including empirical psychology; for (as we saw; cf. chapter II, section 2; chapter III) Kant accepts that psychological states have a causal efficacy which is irreducible to physical causality.

Second, Kant's worry pertains to the *content or validity* that practical judgments purport to have. His explanation of why natural grounds cannot produce an 'ought' judgment is that they can give rise only to a want (Wollen) that "is far from being necessary", whereas the 'ought' "pronounced by reason...forbids and authorizes" our empirical desires. What underlies Kant's point, I think, is the idea that practical judgments purport to have a kind of *normative or rational authority over* empirically given desires. He remarks that the verdicts of practical reason 'confront our desires with a limit and an end' and 'forbid or authorize' acting on those desires. The idea that the verdicts of practical reason have rational authority over our desiderative states means that these verdicts assess and 'rank' desiderative states in terms of distinctively normative concepts: namely, in terms of whether it is *permitted, prohibited, advisable* or *foolish* to act under the motivating influence of those states. To flout such a verdict (e.g., acting in a way deemed impermissible by reason) is to incur a rational deficiency. Now, the idea that the verdicts of practical reason have the kind of authority over desiderative states that allows for such assessment seems hard to square with the idea that these verdicts are caused by, and hence depend on, *those very desiderative states*. That is, if what 'lies behind' or is responsible for a practical verdict is some desire, then it is hard to see how this verdict can claim to have the kind of *status* that seems implied by its ambition to deem desiderative motives required, forbidden, permissible, etc. The fact that a practical verdict is brought about by a desire seems incompatible

²¹² See chapter III, section 3 for a defense of the claim that Kant's notion of spontaneous causality or free action extends to the operations of practical reason with respect to its judgments. Compare GMS, 4: 448: "Now we cannot possibly conceive a reason consciously receiving a bias from any other quarter with respect to its judgments, for then the subject would ascribe the determination of its judgment not to its own reason, but to an impulse. It must regard itself as the author of its principles independent of foreign influences."

with the idea that the verdict presides over desires to assess the rights of the claims that these desires stake. Less metaphorically, what seems implied by the notion of a practical verdict's conferring legitimacy on or rejecting the bidding of desires is a certain kind of *independence* of the judgment from the desires that are the objects of judgment. The naturalistic notion that our practical judgments are caused by, and hence dependent on, our desires implies that there is no such independence. Hence, Kant's point seems to be that natural grounds cannot imbue practical judgments with the kind of authority that these judgments purport to have through employing, in their content, concepts such as (in the case of categorical practical judgments) 'forbidden', 'required' or (in the case of hypothetical practical judgments) 'advisable', 'foolish'.

There is a kinship here between Kant's position and that of philosophers such as G.E. Moore or Warren Quinn. Moore and Quinn point out that there seems to be a gap between the statement that an agent desires an object and the statement that the agent has a reason to pursue the object; the idea that an agent wants, even intensely wants something to happen is perfectly compatible with the idea that the agent has no reason to try to make it happen, or even with the idea that the agent has every reason not to make it happen. The point here is not merely that an agent *may* lack a reason to do x even if she intensely wants to do x, but rather that these two notions are conceptually independent from one another.²¹³ Kant seems to make an analogous point about the lack of connection between the causal history that practical judgments have on the naturalistic picture and the resulting content of those judgments: a mere contingent desire cannot be causally responsible for a valid representation of a normative practical reason.

Surely, Kant is most concerned with *categorical*, namely, moral judgments. We saw (cf. chapter IV, section 1) that those judgments purport to represent a strict normative necessity: they say that certain acts *unconditionally must* be done. With respect to categorical imperatives, Kant's worry is that the claim to strict necessity that such judgments purport to make is incompatible with the *contingent* origin that such judgments would have under the assumption that (as Kant puts it) 'all causality is mere nature'. However, this provokes a general question about Kant's worry: in raising this worry, doesn't Kant commit 'the genetic fallacy', i.e., the (putative) mistake of confusing issues about the causal origin of a judgment (or concept) with issues concerning the judgment's (or concept's) validity?

The question may take an interpretive or a philosophical form. In its interpretive form, it might rest on the firmly entrenched conviction that Kant, among all philosophers, declared questions of origin irrelevant to questions of justification. Since this conviction is especially well-entrenched with respect to Kant's theoretical philosophy, and since analogous worries about the origin of certain theoretical concepts will concern us in the third part of this study, I want to try to dispel the idea that Kant dismisses questions about the origin of theoretical judgments as irrelevant right now, before returning to the topic of the authority of practical judgments.

Here we need to have a brief look at the discussion in the First Critique that precedes the Transcendental Deduction of the categories. It should be uncontroversial that Kant argues that we cannot justify the categories (or judgments that employ them) by investigating their empirical origin (CPR, A 85-7/B 118-20; A 94-5/B 126-8). This point is accepted both by the proponent of the irrelevance of questions of origin to questions of justifications and by my interpretation. The

²¹³ See Moore 1912 and Quinn 1994. Kant puts this point by saying that these two statements relate to different 'orders': the notion that an agent desires to do x relates to the order (or course) of nature, whereas the notion that an agent has reason to do x relates to the 'order of ideas' framed by practical reason; see CPR, A 548/B 576.

difference is that I hold that Kant's Deduction, far from dismissing the relevance of questions of origin altogether, rather seeks to show that the categories have a distinctive *pure* origin and cannot – on pain of losing their claim to objective, necessary validity – be traced to contingent empirical causes. This is shown by the fact that Kant claims, *prior* to the Deduction, that the categories *have yet to show* ("müssen aufzuzeigen haben") a "certificate of birth" other than descent from experience (CPR, A 86-7/B 119), where the notion of a concept's 'descent from experience' here explicitly includes (CPR, B 127-8) 'subjective necessity', the 'blind' production of a concept by non-rational empirical customs of the mind (as in Hume's account). But my point is also supported by the fact that Kant returns to the question of origin toward the end of the B-Deduction (CPR, B 167-8), when he argues that the categories depend for their validity on their independence from contingent, implanted 'subjective necessities'.²¹⁴

In the First Introduction to the Third Critique, we find a passage where Kant makes his stance on the relation between questions of origin and justification unambiguously clear:

[1] If, however, a judgment gives itself out to be universally valid and therefore asserts a claim to necessity, then, whether this professed necessity rests on concepts of the object a priori or on subjective conditions for concepts, which ground them a priori, it would be absurd...to justify it by explaining the origin of the judgment psychologically. [2] For...if the attempted explanation were completely successful it would prove that the judgment could make absolutely no claim to necessity, precisely because its empirical origin can be demonstrated. Now...judgments of reflection...are of the kind mentioned above. They lay claim to necessity and say...that everyone ought so to judge, which is as much as to say that they have an a priori principle for themselves. (EEUK, 20: 239; square brackets added by me)

In this passage, claim [1] states that we cannot justify an a priori concept by appeal to its empirical (psychological) origin. But for Kant, this idea does not follow from the conviction that questions of origin are generally irrelevant to questions of justification: rather, it is justified by the further claim [2] that the empirical origin of a judgment or concept is incompatible with its a priori, necessary validity. And while Kant's line of thought here relates only to judgments that concern a theoretical subject matter, it is perfectly analogous to his claim (in the First Critique) that a necessary practical 'ought' cannot be produced by an empirical want or desire.

Here it is worth clarifying why Kant assumes that the naturalistic picture will posit empirical *desires* as the causes of practical judgments. The point, I think, is that Kant accepts the idea that *practical* judgments are *as such* related in a special way to action. Roughly, to make a practical judgment about what one has reason to do is to be disposed (albeit not, of course, infallibly; maybe only 'insofar as one is rational') to act as the judgment says one ought to act; a deliberator's acceptance of the idea that she should or should not act in a certain way is inseparably bound up with her motivational tendencies to choose and act in the way the judgment commends. Now, for Kant, in the case of moral judgments these motivational tendencies are *pure* (independent of contingent desires) *precisely because* they result from the pure or a priori ideas of reason, which originate in reason itself rather than in on our contingent

²¹⁴ The irrelevance of questions of origin to questions of justification or validity (at least in the case of a priori concepts) is affirmed, for instance, by Bird 1962, pp. 10-1 and Warren 1998, p. 218. But my interpretation is not completely non-standard: in an important article on Kant's philosophical methodology, Henrich 1989, pp. 35-6 rejects the independence of questions of validity from questions of origins when he states that Kant's Deduction "seeks to discover and to examine the real origin of our claim [to a priori knowledge] and with that the source of its legitimacy"; he argues that it is the question of right – not merely the question of fact – that requires a defense of the pure origin of a priori concepts against empiricist challenges.

psychological constitution. But of course Kant was aware of an alternative, empiricist way of accommodating the 'practical nature' of moral judgments. On this empiricist view (the 'desire in, desire out' conception of practical reasoning), both practical reason and the will (understood as the capacity for choice) are the 'slaves of the passions': our practical thought and judgment is (however covertly; see GMS, 4: 419) controlled and produced by the same empirically given desires that are responsible for our choices, and it is the influence of those desires on practical reasoning that accounts for the practical nature of the judgments that result from processes of reasoning. Kant's appeal to the purity or apriority of *both* moral motivation *and* categorical practical judgments, or to the causal independence of both (positively) free acts of choice and free acts of practical legislation, can be understood as a reaction against the empiricist picture.

In the remainder of this section I want to try to show that Kant's intuition that there is a dependence of the validity of categorical practical judgments on their a priori origin is worth taking seriously. Imagine a concrete context in which a putatively categorical prescription is brought up. Suppose A intends to steal money from his sick, disabled father who trusts him, and B tells him that he *must not* do so despite the fact that A strongly desires the money. B claims that everyone has a decisive reason not to act in this manner regardless of what one happens to desire as a matter of contingent psychological fact. Now suppose that the naturalistic conception of the causal origin of practical judgment is correct. If so, what is causally responsible for B's endorsement of her practical verdict that A must not steal from his father are contingent aspects of her psychological disposition, just as A's acceptance of the verdict that stealing from his father is fine is caused by contingent aspects of his psychological disposition. This seems to undercut the claim to unconditional validity that B's judgment purports to make. Generally speaking, the psychological contingency in the origin of purportedly categorical judgments implies that such judgments are causally dependent on *precisely those* factors that the judgments purport to be independent of in proclaiming a practical necessity. B's judgment aspires to give a necessary, desire-independent reason whose normative weight trumps the biddings of a merely contingent desire; this aspiration is invalidated by the fact that her judgment is caused by contingent desires. This intuition is denied rather flatly in the following remarks by Harry Frankfurt:

Rationality and love equally entail selflessness. The differ in that the former is also essentially impersonal. The substance of this difference between rationality and love is not that what a person loves depends largely upon his own personal characteristics, whereas those characteristics play no role in determining what he considers to be required or permitted by reason. The judgments a person makes concerning rationality are manifestly no less dependent than are any other occurrences in his life upon contingent features of his nature and his circumstances. What renders these judgments impersonal is that the claims they make are not limited to the person who makes them; rather, it is implicit that anyone who disagrees with the claims must be mistaken.²¹⁵

Here Frankfurt seems to affirm that my judgment can validly claim that an action is required for *anyone*, regardless of what one happens to care about, even if my reasoning toward and my endorsement of the judgment depends on what I happen to care about. But there is an obvious clash in Frankfurt's scenario between my self-conception as a free thinker who puts forward principles that are intended to be intersubjectively, necessarily binding and my admission that my endorsement of these principles causally depends on features that are peculiar to my subjective, contingent psychology (on 'personal characteristics').

²¹⁵ Frankfurt 1988, p. 90.

Kant's intuition is also rejected in T. Scanlon's conception of "our process of critical reflection": for Scanlon, such critical reflection requires only

...that that process itself be sensitive to reasons, and that later stages of the process be importantly dependent on conclusions reached at earlier stages. But there is no reason, as far as I can see, that this process itself not be a causal product of antecedent events and conditions.²¹⁶

Scanlon rightly stresses that the later stages of a process of practical reasoning must be considered rationally dependent on earlier stages of this process. In the above case, B's judgment that it is wrong to steal from one's father logically depends on (and derives its normative force from) her prior recognition that (say) human persons have a value that is superior to the value of anything that A could buy with the stolen money. But if our reasoning and judging is "the causal product of antecedent events and conditions", then, on a naturalistic view of those events and conditions (and of human reason), our acceptance of ideas that are fundamental to our practical orientation (such as the idea that human beings have an unconditional worth), and our drawing of certain specific conclusions from these ideas, result from the contingent natural constitution of our minds and from the events that lead to the formation of this constitution. This dependence *cannot* be construed in terms of the logical or rational dependency between contents of propositions. Rather, it is a *merely causal* dependence of our reasoning and judging on contingent factors that are, rationally speaking, wholly arbitrary.

Appealing to a case of *conflicting* judgments is a good way of illustrating why the idea that our practical judgments are determined by elements of our contingent psychological constitution seems disturbing. This was the point of my earlier example: if both B's judgment that stealing from one's sick father is prohibited and A's judgment that stealing from one's sick father is fine are determined by B's and A's respective contingent desires, then it is hard to see how B's judgment can (whereas A's judgment cannot) lay claim to having necessary, universal validity. But we do not have to appeal to such conflict here to find the determination of judgment by contingent psychological factors disturbing. As David Lewis puts it:

What comfort would it be if all mankind just *happened* to be disposed alike? Say, because some strange course of cultural evolution happened to be cut short by famine, or because some mutation of the brain never took place? Since our dispositions to value are contingent, they certainly vary when we take *all* of mankind into account, all the inhabitants of all possible worlds. (...) The spectre of relativity within our own world is just a vivid reminder of the contingency of value.²¹⁷

Suppose all mankind judges (i) that human beings deserve respect and (ii) that it is therefore wrong to steal and cheat. If our faculty of practical reasoning and judgment is part and parcel of nature, our acceptance of (i) and (ii) has its origin in natural (e.g., evolutionary) developments. Thus, if the natural history of the human race had been different, we would not attribute a specific value to human beings, or we would find this value compatible with the permissibility of stealing and cheating. But our evaluative outlook in this counterfactual scenario would be the result of the same general types of causes and processes that determine our evaluative outlook in the actual world: in both scenarios, our respective acceptance or lack of acceptance of (i) and (ii) is a consequence of our minds being contingently 'wired' in a certain way. I think that on this naturalistic conception of our mind, the idea that our acceptance of (i) and (ii) is "sensitive to

²¹⁶ Scanlon 2003, p. 370.

²¹⁷ Lewis 2000, p. 89.

reasons", or the idea that (i) and (ii) express an objective rational necessity, seems rather elusive. Practical reason here is the slave of our contingent psychology or neurology: the 'output' of our reasoning procedures is crucially determined by the 'input' we happen to have received during the processes that led to the formation of our minds, brains or characters.

At least on the face of it, this line of thought casts a doubt *only* on the possibility of *categorical* practical judgments. Suppose that the natural history of the human race had equipped us with an intense appetite for insects. Under the assumption that normative practical reasons depend on contingent desires, this might show precisely that we ought to eat insects. Subjective, desire-dependent reasons are *supposed to be* rationally arbitrary and contingent, so there seems to be no problem in conceding that our reasoning toward the judgment that we have such reasons (e.g., that we ought to eat insects) is equally contingent upon rationally arbitrary events. The contingency of the judgment's validity corresponds to the contingency of its causes. But this correspondence breaks down in the case of *categorical* practical judgments: here the contingency of the factors that determine us to make the judgment clashes with the judgment's claim to necessary validity. The fact that subjective and objective practical judgments have – on the naturalistic picture – the very same contingent causal ancestry tells against the idea that we are dealing with two distinct classes of practical judgments (and reasons).

Now, one may insist that the reason why we can and should bracket questions of causal origin with respect to questions of validity is that the question of validity is one of *truth*, and assessing the truth or falsity of a judgment and explaining its causal origin are wholly different forms of enquiry. But that there is no complete independence between these two forms of enquiry can be shown rather easily: in a case of optical illusion, the falsity of my judgment that (say) the stick in the water is bent is obviously connected to the fact that my judgment has its origin in (is explained by) some internal mechanism *other than* veridical perception of an objective property. Of course, this is a special case: our perceptual judgments are generally veridical, and what this suggests is that generally speaking questions about origin do not show up in questions about the truth of judgments because it is implicitly assumed that those judgments result from more or less direct (perceptual) contact with their objects. And this raises a point that is important for the purposes of our discussion of a putative link between the origin and the validity of categorical practical judgments. Surely, a naturalistic conception of human reason could be shown to be compatible with the possibility of categorical practical judgments if it could be held that necessary values and reasons are an object of *perception*. For in this case one could argue that the natural history of the human mind or brain equipped us to track down necessary values and reasons as objective features of reality. But this perceptual account of practical judgment faces well-known metaphysical and epistemological problems. The metaphysical idea that external reality contains values that radiate necessary prescriptive force is hard to understand; and it is equally hard to substantiate the claim that we could claim to know these entities on the basis of some kind of perceptual or intuitive access. In any case, it is very unlikely that someone committed to a naturalistic or empiricist outlook would posit such entities or perceptual capacities²¹⁸: remember that the naturalist we are concerned with allows only

²¹⁸ There are naturalists, such as Boyd 1988 or Railton 1986, who argue that moral facts are identical to or constituted by natural facts and hence perceivable. But even if such reductive moral realism were generally acceptable, it would not account for the possibility of literally perceiving the *unconditional prescriptive force* of moral facts. Boyd, Brink and Railton are externalists about moral judgments: they think that recognizing the *truth or validity* of moral claims does not require the acceptance of the *authority* (the prescriptive force) of such

powers and causes that figure in the explanations of natural science (cf. p. iii in my Introduction). Thus, while there is a tight and obvious link between the causal origin and the truth of (say) judgments about the shapes of objects, there is an equally obvious gap, on the naturalistic picture, between the causal origin of judgments that state that certain acts unconditionally must be done and their purported objective, necessary validity. To explain why we make those latter judgments, we cannot appeal to a correspondence between the content of those judgments and perceivable features of external reality. So we must turn our gaze inwards – and on a naturalistic conception of human reason, all we find here are psychological or neurological processes, flowing from an empirical constitution that developed in the way it did because certain genes happened to mutate or (on a smaller scale) because society happened to condition its members in certain ways. There appears to be a misfit here between (on the one hand) the utter contingency in the origin of our evaluative attitudes, i.e., their dependency on rationally arbitrary empirical conditions, and (on the other hand) the claim to rational necessity inherent in the idea that (say) certain kinds of actions are such that someone who performs them incurs a rational deficiency.

Let us now try to clarify the precise form of Kant's argument. I suggest the following schema: (I) A practical judgment can validly represent an 'ought' only if it has the normative authority that is required for deeming courses of action permissible, obligatory, wrong, foolish, or advisable. (II) If a practical judgment is determined by natural causes such as empirical desires, it lacks normative authority. (III) All practical judgments are determined by natural causes such as empirical desires. (IV) (from (II), (III)) All practical judgments lack normative authority. (V) (from (I), (IV)) No practical judgment can validly represent an 'ought'.

One might object to the general spirit to this line of argument; for instance, one might hold that it derives a specious kind of plausibility from a slide between concerns about the content or validity of normative practical judgments and concerns about the 'authority' of a judge who puts forward normative verdicts. Of course, for Kant there is no such slide here: for him the idea that we make judgments that have a special normative content (that express rational necessities) is inseparably bound up with our self-conception as judges who have a special authority by virtue of the independence of our faculty of judgment from contingent empirical conditions (GMS, 4: 448). Since we are the ones who are responsible for those judgments, we cannot think about those judgments in abstraction from our self-image; whatever can correctly be said about us as subjects of reasoning has implications for the content of the judgments that issue from our processes of reasoning. And if we adopt the naturalistic image of man and thereby concede that our verdicts about what there is reason to do are determined by our contingent empirical make-up, then it seems hard to give substance to the idea that the practical judgments that flow from this make-up have a validity or normative force that is independent of contingent empirical conditions. As I pointed out earlier, this line of thought could be blocked if we could defend the idea that practical necessities are an object of (quasi-)perceptual access, for in this

claims. By contrast, Mackie 1991 supposes that recognizing a moral truth would have to entail a recognition of its categorical authority; and he argues that all moral judgments are false *precisely because* he thinks that these judgments could be true only if there were 'queer' entities whose necessary prescriptive force we could 'intuit'. Whatever one thinks about the precise contours of Mackie's internalism, and quite regardless of the many difficulties that afflict the externalist-reductivist model, it is clear that this model does not provide a response to the worry that animates Kant (and Mackie, and Lewis): namely, the worry about how we can make sense of the idea that certain practical judgments have necessary prescriptive force if those judgements are caused by contingent elements in our psychological constitution.

case the special content of our practical judgment is guaranteed by something external to us and thus unaffected by the contingency of our internal constitution; but this perceptual account faces considerable difficulties (some of which I indicated above).

Before discussing the soundness of this argument, a potential ambiguity in Kant's position must be noticed. It is not crystal clear whether Kant's worry is that (i) on a naturalistic conception of human reason, all our practical judgments fail to represent normative reasons, or (ii) that they falsely represent such reasons. If we take him strictly at his word (CPR, A 547-8/B 575-6), we must read him as saying that contingent natural states such as empirical desires cannot even give rise to the representation of a practical 'ought'; but he also suggests (GMS, 4: 419) that the "so-called" categorical imperative would falsely appear as having unconditional authority if it was conditioned by contingent desires. My formulation of the argument, as being concerned with the possibility of *valid* representation of practical 'ought', is more closely aligned with this second, weaker alternative. In this chapter, will confine myself to discussing the worry that all practical ought judgments would be false on a naturalistic conception of human reason, although I think that a concern about the *meaning* of practical judgments can be motivated.²¹⁹

Now, regarding the premises of the above argument, a naturalist is clearly committed to (III). This leaves two substantive premises. (I) should be rather uncontroversial: practical 'ought' judgments prescribe what one should do, and Kant is clearly right to emphasize that inherent in such prescriptions is a claim to what I call 'normative authority', i.e., an appeal to a standard of practical rationality whose violation implies a rational deficiency that calls for criticism. So the soundness of the above argument hinges on the cogency of (II).

(II) claims that if the causal ground of a practical judgment is a contingent desire, the judgment lacks normative authority (i.e., fails to validly represent a normative practical reason). In assessing this claim, we need to differentiate between categorical and hypothetical judgments. I have already argued (in this section) that there is something to be said for Kant's idea that the claim to universal, necessary validity made by categorical practical judgments is incompatible with the concession that these judgments are determined by contingent empirical desires. For present purposes, I want to refrain from discussing this point any further (I shall briefly return to it in section 4), because I do not think that it is *all that* controversial. Many naturalists would *agree* that a naturalistic outlook is incompatible with the idea of there being (correct) categorical imperatives, precisely because the contingency of our evaluative attitudes nullifies their claim to universal or necessary, objective validity (cf. the Lewis quote on p. 95 above). But they would surely balk at the idea that a rejection of categorical imperatives implies the impossibility of practical normativity *as such*. They would argue that the determination of our practical orientation and evaluative attitudes by contingent desires is perfectly compatible with there being meaningful practical prescriptions *that derive their authority from our desires*. As Bernard Williams says: "Desiring to do something is of course a reason for doing it."²²⁰

²¹⁹ I have tried to spell out what underlies this concern above on pp. 88-9 (see also footnote 210). In a nutshell, it seems to me that the notion of a false practical judgment implies a kind of negative significance that requires (in analogy with the idea of bad action) that we have rational faculties that are something over and above the 'play' of empirical forces and processes: it implies that reason erroneously endorses the bidding of a desire. It should also be noted that Kant's idea that all practical oughts are meaningless if one has only the course of nature in view pertains not merely to the content but also to the *formula* of practical judgments (for this distinction, see chapter IV); I will discuss this concern (which is quite different from the present one) in chapter VI.

²²⁰ Williams 1986a, p. 19.

The dialectical problem here is that as long as we focus on the issue of categorical imperatives, we face the threat of a stalemate. Kant would insist that unless we accept that our practical reasoning and judging is not exclusively conditioned and determined by empirical desires, we cannot account for the possibility of categorical imperatives; but the naturalist would simply respond that she does not believe in categorical imperatives anyway. Kant would then point out that giving up on categorical imperatives implies giving up on a meaningful moral practice, but many naturalists *already* accept that our moral practice is confused in its reliance on valid categorical imperatives.²²¹ So an appeal to the incompatibility of a naturalistic conception of practical reason and judgment with the possibility of categorical imperatives is unlikely to impress a naturalist about normativity.

²²¹ We already noticed Mackie's view that all our ordinary moral judgments false because there are no categorical imperatives, or unconditional values; this is echoed in Joyce's 2007 view that the reliance of our morality on categorical imperatives motivates a 'fictionalist' view of morality. It is also echoed in Lewis' dispositional theory of value: Lewis holds that "strictly speaking, Mackie is right: genuine values would have to meet an impossible condition" (2000, p. 93), and he suggests (*ibid.*, pp. 91-2) that our intuitive resistance to the notion that values are contingent ("It feels wrong") is due to the (irrational, it is implied) strictures of moral discourse ("because a large and memorable part of our discussion of values consists of browbeating and being browbeaten"). In a very similar vein, Williams 1995 claims that our practice of holding people responsible for violating moral rules is confused – not because of its commitment to the falsity of determinism, but because it requires the *fiction* that those people had reason to act morally when in fact they had no such reason. Williams position is related to other Neo-Aristotelian attacks on the notion of moral obligation, such as that of Anscombe 1958 and of Foot 1972. Finally, Harman 2000 argues that the fact that (e.g.) an assassin has an 'informed desire' to shoot his victim is sufficient to exempt the assassin from the moral requirement not to kill. This, of course, leaves wholly obscure what the point of having such a 'requirement' would be in the first place.

An exception here is Schroeder 2007, who thinks that a desire-based theory of practical reasons is compatible with the ordinary conviction that moral requirements are reason-giving for every agent. The basic idea seems to be (p. 109) that a single reason can be 'overdetermined', i.e., by explained by appeal to many different desires, and (pp. 114-5) that moral reasons are such that they can be explained by *any* given desire, i.e., that having any desire whatsoever implies a reason to comply with moral demands. The problem is that Schroeder does not at all explain these bold ideas. His excuse is that "this broad-brush defense of Hypotheticalism is...not the place to become bogged-down in the details of such an explanation". But without any such details his ideas lack philosophical substance. The following passage (*ibid.*, p. 142) is revealing here: "If Ryan can't stand Katie, Ryan may have abundant reasons to place less weight on this reason [to help Katie]. But those reasons aren't relevant to its weight, because they won't be of the right kind. A reason has a certain weight just in case it is correct to place that much weight on it. And correctness is determined by reasons of the right kind. (...) that means that they must be reasons that everyone who is placing weight on reasons has, in virtue of being someone who is placing weight on reasons. But the activity of placing weight on reasons is just the activity of deciding what to do. So it is simply the activity that every agent is engaged in. So the right kind of reasons with respect to the correctness of placing weight on reasons are precisely the class of agent-neutral reasons. It follows that Ryan's idiosyncratic reasons to place less weight on his reason to help Katie are irrelevant, the wrong kind of reason to determine its weight." This has the form of an argument ("so...so...it follows") but I find it impossible to discover the logical connection between the individual sentences. Price 2009 comments that this passage is a mere display of "verbal prestidigitation" and that Schroeder's "only expedient is to cast a veil of words over the vacuity". This is harsh, but justified. The natural question is: how is it supposed to follow from the fact that Ryan is engaged in 'the activity of placing weight on reasons' that he has a reason to help Katie, even though he has a strong desire to see Katie suffer and *even though* (given Schroeder's desire-based theory of reasons) Ryan's reasons are explained by his contingent desires? Schroeder does not address this question in an upfront manner, and he sees no need to provide a defense of the intriguing but elusive suggestion that everyone who engages in the 'activity of placing weight on reasons' has a reason to be moral. Without such a defense, his suggestion remains a mere, empty promise. I further discuss Schroeder's account below (toward the end of section 2).

For this reason, I will not argue for (II) by appealing to the requirements of categorical imperatives. Rather, I will argue that the naturalistic attempt to reject (II) on the basis of positing an intrinsic link between normative reasons and desires raises problems that should worry even those who reject categorical imperatives but who still believe in practical normativity.

2: Desire-based theories of normative practical reasons

(II) states that if a practical judgment is determined by contingent empirical desires, it fails to validly represent a normative practical reason. But if there is a tight link between what someone desires and what her normative practical reasons are, the idea that our practical judgments are caused by contingent desires is rendered harmless. Naturalists about normativity typically argue that appealing to naturalistically acceptable resources such as desire of various kinds and orders is sufficient to account for the validity of normative judgments. They might claim that a statement such as, 'A has reason to do x' can be *analytically reduced* to a statement such as, 'A desires to do x': *what we mean* when we say that A has reason to do x is that A has a certain kind of desire to do x. The problem with this approach is that it invites Moore's open question argument; and thus contemporary naturalists about practical normativity tend to be indifferent about the possibility of an analytic reduction and instead focus their attention on the idea that *metaphysically speaking*, normative facts concerning what an agent has reason to do depend on facts about what the agent desires.²²² If this theory were correct, then a practical judgment's being the effect of some desiderative state cannot tell against the judgment's having normative authority, because the very state that is the causal source of the judgment also fixes the normative fact that the judgment seeks to represent. The 'want' can produce an 'ought' since, metaphysically speaking, practical 'oughts' rest on 'wants'.

Here an important clarification is in order. It seems that desire-based theories of practical reason concern the *truth-conditions* of normative practical judgments. So why can't a desire-based theory be completely neutral about the *causal origin* of practical judgments? Notice, first, that certain questions of causal origin *are* essential to desire-based theories: proponents of such theories hold that their theory is supported by the idea that a consideration can be a reason for Tom only if it can actually move Tom to act and that all actions are motivated (i.e., caused and explained) by desires.²²³ Now, notice that this idea does not by itself establish that all of Tom's reasons depend on his desires: for one might concede this idea and argue that Tom can acquire a desire to do x *upon recognizing* or judging that he has reason to do x, in which case the existence of the reason (the object of Tom's recognition) might be prior to, and independent of, Tom's desires. Hence, it is essential for desire-based theorists to hold that a deliberator's desires cause and explain not only her actions but also her deliberation and judgments about what her reasons are; for only thus do we get a dependency of *practical reasoning* on empirically given desires.²²⁴

²²² This strategy of evading Moore's point – shifting the focus from conceptual to metaphysical analysis – is recommended by Harman 1977, pp. 19-20 and by Schroeder 2007, pp. 60-5. Schroeder seeks to give a reductive metaphysical (rather than semantic or epistemic) analysis of normative practical reasons to desires.

²²³ See Williams 1997 and Harman 2000. Compare Milgram 1997, pp. 4-5, speaking in the voice of the desire-based theorist (not his own): "How could anything be a reason for action if it could not motivate you to actually *do* something? And what could motivate you to do something except one of your desires?"

²²⁴ This line of thought is particularly perspicuous in Williams 1997, who argues (correctly, in my view) that the decisive issue between the desire-based theorist ('internalist') and his 'externalist' opponent concerns the question of "what it is to come to believe" a statement about what one has reason to do (p. 367). This is a psychological

Under the plausible assumption that practical reasons can be recognized only through reasoning, and in this sense depend on reasoning²²⁵, the dependency of practical reasoning and its output (practical judgment) on contingent desires (i.e., on the deliberator's 'subjective motivational set') suggests that practical reasons themselves depend on contingent desires.

The accuracy of this way of characterizing the commitments of desire-based theories is, it seems to me, confirmed by the fact that it leads directly to the Humean 'desire in, desire out' conception of practical deliberation that is standardly associated with desire-based theories. On this conception, fundamental psychological dispositions yield the 'input' of practical deliberation and thereby control its output – judgments about what one should do as well as the newly acquired motives that underlie those judgments. The naturalistic allure of the desire-based theory really emerges when it is understood *both* as a theory of reasons *and* as an account of practical reasoning: the desire-based theory demystifies what it is to recognize a reason and to come to be accordingly motivated by picturing this as a psychological process that begins from antecedently given, general desires (such as a concern for the well-being of one's friends, or thirst) that (in conjunction with beliefs about non-psychological facts) give rise to more specific ideas about what to do (e.g., the idea that one should call up one's depressed friend, or that one should put a dollar into the soda machine). Hence, desire-based theories are normative as well as psychological: their proponents typically argue *from* the explanatory potential of their account *to* its adequacy as a normative theory.

But what about third-person judgments? I think that for the desire-based theorist, such judgments, if veridical, are also, albeit in a different way, caused by the desires that make the judgment true. Of course a third-person judge need not *have* Tom's desires to judge what Tom has reason to do, but her judgments are (if correct) caused by Tom's desires insofar as they result from a (however indirect) perceptual recognition of those desires. Here, again, the naturalistic allure of the desire-based theory is that it renders the perceptual recognition of a normative fact entirely unmysterious: there is no need to explain how prescriptive or normative force can be an object of perceptual recognition, precisely because, on a desire-based theory, third-person judgments about reasons that originate from the perception of someone else's desires *have no prescriptive or normative force for the perceiver or judge*. Consider an assassin who judges that his victim has a reason to prevent him from killing her. This judgment is (on a desire-based theory) caused by the assassin's (however indirect) perception of the victim's desire to live, just like his judgment that there is a chair in the room is caused by her perception of the chair. Neither judgment has (on the face of it) any practical or normative implications for the assassin.

Thus, it is an important aspect of a desire-based theory that the truth-conditions of judgments about reasons coincide, in one way or another, with the causes or explanatory conditions of judgments about reasons. On such theories, there is a neat correspondence between the contingency of our evaluative thoughts and the contingency of the values represented by these thoughts. Both have, ultimately, the same source: our empirically given, contingent psychology. If such theories are correct, the dependency of our practical thought on contingent, rationally arbitrary conditions cannot count against the validity of our practical thought, because

question; Williams argues that his externalist opponent cannot answer it because the only way of answering it collapses into internalism (p. 368). For Williams, the answer is (p. 365) that a conclusion of practical reasoning, a judgment that one should do x because doing so satisfies some element of one's 'S' (subjective motivational set), is "of course controlled by other elements in S, if not necessarily in a very clear or determinate way".

²²⁵ See Harman 2000, p. 30 ("reasons seem to depend on reasoning").

the very objects of such thought (our reasons and values) depend on those same conditions. Thus, considering the cogency of such theories is of utmost importance for assessing Kant's argument that contingent desires cannot produce valid practical 'oughts'.

We can start our examination of desire-based theories by noting that the desire-based theorist must substantiate the idea that we can grant normative authority to some desires rather than others – that is, she must single out some desires from the overall set of an agent's desires as those that explain what the agent ought to do. More precisely, the desire-based theorist needs to single out those *properties* of desires which can function as a reliable guide toward what an agent really wants and by reference to which we can specify what the agent should do. The need to single out some desires among others that possess normative authority can be motivated in various ways. First, an agent can be subject to conflicting desires at a given point in time: it can be the case that A desires to both x and to y, where x and y stand for actions that are mutually exclusive so that x-ing implies not-y-ing and vice versa. Here we must say that at least one of those conflicting desires is not worth acting upon, for if we cannot say this we will end up with the notion that A both should and should not do x and both should and should not do y.

Second, a plausible account of practical normativity must entitle us to combine the claims (i) that an agent A desires to do x and (ii) that A should not do x. A theory on which an agent's desiring to do x implies that she should do x, or on which every desire as such has conclusive normative authority, would face a number of problems. It would deprive practical reasoning of its point; for if all desires give us the same reasons for acting, we do what we should do whenever we act on the basis of some desire, and thus no reasoned judgment is needed to tell us what we should do. Moreover, such a theory would yield absurd results in cases where a desire is based on false beliefs, such as Bernard Williams' example where someone has an intense desire for gin and falsely believes that some glass, that actually contains petrol, contains gin. Finally, we can only make room for forms of criticism of an agent's practical rationality (e.g., by employing the notion of weakness of will) if it is not the case that whenever the agent desires to do x it follows that she ought do to x.

For these reasons, I shall presuppose that it is a condition of the adequacy of desire-based accounts that they can distinguish between desiderative states that have and desiderative states that lack normative authority, and that they are not forced to the conclusion that whenever an agent acts on some desire it follows that she has reason to act in this manner.

This condition cannot be satisfied by appealing to the strength of an agent's desires. The strongest desire to which an agent is subject at a time is the one that actually moves the agent to act. Hence, if the desire-based theorist were to single out current strength as the property which distinguishes reason-giving desires from desires that lack normative authority, then she would end up with the position that every agent at every time does what he has reason to do and that there is no way of drawing a distinction between what an agent does and what she should do. The notion of a normative reason would then collapse into the notion of a motivating reason, and this would amount to an abolition of practical normativity: that is, it would imply that there is no notion of normative practical guidance that can be contrasted with a notion of psychological prevalence or efficacy.

This is why desire-based theorists characteristically try to account for the normative authority of some desires rather than others by appeal to counterfactual rather than actual strength. This brings us to the notion of 'informed desire' which plays a crucial role in desire-

based accounts of practical normativity.²²⁶ The idea is, roughly speaking, that we should grant normative authority to those desires an agent *would have* if she were to go through a process of *informed deliberation*.²²⁷ Informed deliberation is not meant to be deliberation which makes the agent aware of facts that give him reason independently of what his desiderative states are. This would be incompatible with desire-based theories. The process of informed deliberation is meant to highlight those elements of the agent's psychological constitution (her 'subjective motivational set') that are truly reason-giving for the agent. It is part of such a process that the agent would find out all the relevant facts; e.g., in the above example, the agent would learn that the glass on the table contains petrol instead of gin, and would then lose her desire to drink the stuff in the glass. Another feature of the process of informed deliberation is that the deliberator exercises her imagination in such a way that she no longer suffers from misconceptions about the effect of certain events on her psychology. On this basis, the desire-based theorist hopes to substantiate the claim that we have no reason to choose some lesser near-future good over a greater far-future good even though we have a prevalent desire for the lesser near-future good. The greater spatiotemporal distance to future events reduces the vividness of our imagining of what it would be like for us to suffer or enjoy the relevant consequences; conversely, the greater proximity of an event that lies in the near future increases the vividness of our imagining what it will be like to be in pain or to feel pleasure. Informed deliberation is meant to correct this bias via an exercise of the imagination which adequately represents the amount of future pain or pleasure so that we become motivated to choose what is objectively the greater amount of pleasure or the lesser amount of pain. It seems clear that false representations of the effects that future events have on our psychology abound and that they are a typical target of practical criticism: for instance, think of the tendency to avoid going to the dentist in full knowledge that this is likely to lead to a long period of great pain some decades down the road.

The claim here is *not* that A would choose to avoid future pain if she became aware, through the right kind of deliberation, what a terrible thing future pain is. Rather, the idea is that an aversion against future pain is *already* part of A's subjective motivational set before the process of deliberation sets in. This aversion 'lies dormant', so to speak, in a way that is somewhat analogous to the way in which the agent's aversion against drinking petrol lies dormant in Williams' example. In this latter case, A has a false factual belief about what the glass contains. As soon as that belief is replaced with the belief that the glass contains petrol, the aversion against drinking petrol exercises motivational force. By analogy, the reason why the aversion against future pain lies dormant is that the agent's imagination *falsely represents* a future event. Due to the temporal distance of the future events which would cause the pain, the imagination does not accurately represent the event as it is going to affect the agent; it pictures the event in a much attenuated form. Informed deliberation would invigorate that representation and bring it closer to the truth; the lively coloring of that invigorated representation would then trigger the aversion against future pain and render this aversion motivationally efficacious.

The appeal to the notion of informed deliberation promises to give us the resources for singling out those members of an agent's set of desires that have normative authority over others, in the sense that we can say about an informed desire but not about uninformed desires that it

²²⁶ See Brandt 1979; Harman 2000; Williams 1997.

²²⁷ Brandt 1979 appeals to what agents would desire after a process of cognitive psychotherapy, but this raises many problems; chiefly, the problem is that most of us lack any desire to expose ourselves to such a process. Therefore I shall stick with the appeal to informed deliberation in what follows.

gives rise to or constitutes a reason to act. We now need to examine whether the notions of informed deliberation and informed desire live up to this promise.

Suppose that S is a person whose psychological constitution does not contain 'eccentric' desires such as a longing to be in a state of pain. S has an ardent desire to cheat on his pregnant wife whose rich father is his employer. S knows that acting on this desire will give him immediate short-term pleasure, and S also knows that his act of adultery will cause him to be in a long state of utter misery and despair. Now suppose further that S knows (inductively) that if he does make a certain imaginative effort to conceive what it is like to be in utter misery, his desire for future happiness (fueled by the vivid prospect of future misery) will grow stronger and will come to exceed the force of the desire to cheat so that he will be motivated to act in accordance with his desire for future happiness. And suppose further that this knowledge does not suffice to move S to exercise his imagination to get rid of the desire to cheat. Here it will be difficult for the desire-based theorist to substantiate the claim that S has reason not to cheat on his wife. The desire-based theorist's envisaged strategy for assigning normative authority to the agent's desire for future well-being, despite the fact that this desire is not actually motivationally efficacious, consists in the appeal to the true counterfactual that the agent would be motivated to act on his desire for future well-being if he went through a process of informed deliberation. But S has no motivationally efficacious desire to engage in a process of informed deliberation. So the only way to exploit the envisaged strategy for the purpose of granting normative authority to *this* desire is by saying: *S would be motivated to choose to engage in informed deliberation if he went through a process of informed deliberation.* This seems viciously circular.

Why, precisely, does this raise a problem for desire-based theories? The appeal to informed deliberation was invoked to avoid the conclusion that a 'capricious' desire can provide practical reasons. In fact, the notion of informed deliberation (via the related notion of counterfactual strength) was supposed to make room for the very possibility of drawing a distinction between whimsical desires on the one hand and desires that reflect the agent's real ends on the other hand. So if we want to distinguish between merely whimsical and normatively significant desires when the intentional object of the desire is 'engaging in the activity of informed deliberation', we need a criterion that is independent of the intertwined notions of informed deliberation and counterfactual strength in order to settle the question of whether an agent should engage in informed deliberation. If there is no such independent criterion, we cannot draw the distinction, and this means that the desire-based theorist must take an agent's present desire to engage or not to engage in informed deliberation at face value – that is, as indicating what the agent has reason to do.

Could the desire-based theorist insist that S should engage in informed deliberation even if S lacks a motivationally efficacious desire to do so? To the extent that S has *some* desire to engage in informed deliberation, it follows, for desire-based theorists, that S has *some* reason to engage in informed deliberation.²²⁸ But this cannot show that S has a conclusive reason to or that he *should* engage in informed deliberation: i.e., that the desire to engage in informed deliberation has normative authority. The problem is that the appeal to informed deliberation was supposed to give us the distinction between whimsical and authoritative desires in the first place. If this criterion cannot be applied to a desire, there are no grounds left for normatively discarding the

²²⁸ Schroeder seems to think that any desire whatsoever gives a (however weak or 'non-weighty') reason; see his 2007, p. 90 ff. I discuss his account in more detail below.

desire. But this criterion cannot, on pain of incoherence, be applied to a pro/con-attitude which has 'engaging in a process of informed deliberation' as its intentional object. So the desire-based theorist cannot normatively discard an aversion against engaging in informed deliberation, *unless* she can give us a criterion for determining what an agent has reason to do that is independent of the appeal to informed deliberation. But as a desire-based theorist, she is constrained by the condition that this criterion can *only* relate to properties of given desires.

This raises a general worry about whether it is possible to settle the question of what an agent has reason to do solely on the basis of *counterfactuals* about what the agent would desire to do under some non-actual condition. It seems that, on a desire-based theory, no mere appeal to counterfactuals is going to yield a criterion for specifying what an agent really wants and what she has reason to do unless we can hold fixed some *actual* aspects of her psychological constitution. We can see this by pointing out that there may be all kinds of counterfactuals that yield truths about what someone would desire in some counterfactual scenario. Suppose S is a cynical misanthrope whose greatest joy in life results from contemplating the failures and miseries of his fellow human beings. We may correctly say about S that if S were subjected to something like the Ludovico Treatment (cf. Clockwork Orange), S would lose his desire to seek out and rejoice over instances of human misery and failure. This counterfactual is true because there is a causal process which, if initiated, would lead from S's current misanthropism to a philanthropic outlook. No desire-based theorist would want to say that this counterfactual specifies what S should do, and this is because S's attitude toward the process specified in the counterfactual is one of hostility, not of acceptance. But *that* his attitude is one of hostility is borne out by his *current* desires. We can rule out the possibility that S *really* desires to be a philanthrope – despite the fact that there is a causal route from his present motivational states to a philanthropic outlook – only by taking at face value his present, prevalent aversion against philanthrophism. An appeal to informed deliberation can thus succeed as a criterion for distinguishing between S's real interests and his merely whimsical desires only under the condition that S *accepts* this criterion; and by acceptance, we can mean nothing less than *actual and motivationally efficient* acceptance. In order for it to be true that S accepts the criterion of informed deliberation as a guide toward what he really wants (i.e., as a guide to those desiderative states that yield reasons for acting), it has to be the case that S actually does strive to deliberate in an informed manner and hence to exercise his imagination in ways that will cause him to lose desires such as (in the above example) the inclination to cheat on his wife.

The desire-based theorist might respond here that it is implausible to assume that an agent whose psychological constitution contains an aversion against future pain would not be moved to engage in a process of informed deliberation by her recognition that engaging in informed deliberation would cause her to lose the desire to act in a way that leads to future pain. But I do not think that it can be a strict condition of the ascription of a desire to avoid (lasting, intense) future pain that an agent engages in informed deliberation upon learning that this would cause her to act in ways that would avoid such pain. The very same 'whimsical' desire that causes people to discount their long-term happiness might also cause them to fail to engage in informed deliberation, or at least to fail to exercise their imagination in such a way that their prevalent 'imprudent' desires are indeed changed as a result of the deliberative process. Suppose, again, that A is an average agent (no masochist etc.); A has to choose between a minor amount of near-future pain and a long period of far-future pain; due to the vividness of the prospect of the

temporally close near-future pain and the faintness of the prospect of the distant far-future pain, A feels driven to choose to avoid near-future pain. Now suppose A knows that if he were to engage in informed deliberation, he would acquire the desire to avoid far-future pain and lose the desire to avoid near-future pain. That knowledge by itself need not have any effect on A. Given that A is *already* in the grip of the vividness of the prospect of near-future pain, it is possible that knowledge of the counterfactual would not be sufficient to get him out of that grip. What *would* get A out of that grip is not the *knowledge* of the truth of the counterfactual but *going through* the process of imagination which figures in the content of the antecedent of that counterfactual. But it seems possible that the vivid prospect of near-future pain, and the resulting desire to avoid near-future pain, causes A to be unable to engage in informed deliberation or at least prevents A's imaginative effort from having an impact on his psychology that is strong enough for him to lose the 'imprudent' desire. So, the same whimsical desire that leads A to choose what – intuitively – he has no reason to choose might also lead A to choose not to engage in informed deliberation, or might prevent the process of deliberation from preempting the force of the 'whimsical' desire.

The desire-based theorist could respond by biting the bullet, i.e., by saying that *since* A's and S's desires prevent them from doing what it takes to avoid acting on their putatively whimsical desires, it follows that the desires are *not* whimsical but rather disclose what A and S really want, so that A *has* reason to choose a minor near-future pain over a long period of far-future pain and that S *has* reason to cheat on his wife. The desire-based theorist could add that this response seems problematic only to the extent that one presupposes desire-independent values, for instance that there is something intrinsically bad about (say) being in a long state of pain. If my argument did rest on this presupposition, it would beg the question against desire-based accounts.²²⁹ But the argument makes no such presupposition. We can concede to the desire-based theorist, for the sake of the argument, that (say) a masochist may have no reason to choose to avoid future pain. I have argued from the premise that there is a psychologically real distinction between persons with an aberrant (e.g., masochistic) psychological constitution and average people who are led to choose a great amount of far-future pain not on the basis of some psychologically rare fetish, but on the basis of what I have called a whimsical desire that is in no way representative of their overall evaluative outlook. It would be a gross sacrifice of the phenomena to assimilate such people to persons that have a rather uncommon (e.g., masochistic) psychology. Since desire-based theories are committed to the view that practical reasons are

²²⁹ This, I think, is why many of the arguments that Parfit 2011 adduces against desire-based theories seem a bit pointless. For instance, he tries to show that desire-based theories are committed to the absurd notion that someone who desires to be in state of agony has reason to get herself in such a state. But the notion that a person may have a reason to get herself into a state of agony is absurd *only if* pain is, as Parfit puts it, "bad by its nature" – and the desire-based theories does not accept that any state is intrinsically good or bad independently of what a person happens to desire. Similarly, Parfit's claim that desire-based theories imply that "nothing matters" cuts no ice: the desire-based theorist already accepts that 'nothing matters' in the sense that there are not (as Parfit calls them) "object-given reasons" for doing things that do not depend on what people happen to care about. The desire-based theorist is likely to add that as long as people do care about things, they do not need 'object-given reasons', and to the extent that they do not care, 'object-given reasons' have no bearing on their deliberation. For the same reason, I think that the arguments against desire-based theories advanced by Shafer-Landau 2003 and Smith 1994 have limited force: they point out that desire-based theories imply that normative discourse would be profoundly relativistic, but this is precisely the point that such theories are intended to have: their foundational intuition is that it would be a mistake to assume that there are practical reasons that apply to anyone. Here we seem to end up with a stalemate similar to the one we ran into at the end of section 1.

rooted in the psychological constitution of agents, it follows that if they held that the masochist on the one hand and the ordinary (weak-willed) agent on the other hand have the same practical reasons, they would be committed to the view that these two agents have a similar psychological constitution. And this seems like a *reductio*.

Let us now consider whether the difficulty I have belabored also shows that the appeal to informed deliberation fails to avoid the result that according to desire-based theories an agent has reason to do *whatever she ends up doing*. We saw that the desire-based theorist is committed to holding that the appeal to informed deliberation can only function as a standard for what an agent really wants and has reason to do if the agent is actually motivated to engage in informed deliberation; or, generally speaking, that *no* correct *counterfactual* specification of what an agent would desire under some non-actual condition suffices to show that the agent has reason to act on the counterfactual rather than on his actual desires. But if the appeal to an agent's counterfactual motivation is insufficient to show what the agent has reason to do, then it seems that we are indeed forced – if we accept a desire-based theory – to appeal to the agent's actual prevalent motivation in order to determine what she has reason to do. That is, we end up with the idea that the specification of what an agent has reason to do requires, essentially, a reference to her currently prevailing desire. The appeal to the notion of informed deliberation was surely intended to deliver a *generic* procedure for determining what a given agent has reason to do quite irrespectively of her actual desires, but it seems as if the appeal makes a presupposition that desire-based theorists are not entitled to, namely, the presupposition that there is a normative reason (to engage in informed deliberation) that is not constituted by and hence does not reduce to facts about desires.²³⁰

A good test for the seriousness of this difficulty is Williams' case of the man, D, who desires to drink from a glass that, unbeknownst to him, contains petrol: here it seems *just obvious* that D has no reason to drink from the glass even though he is actually motivated to do so. The desire-based theorist must attempt to accommodate this intuition by appealing to the counterfactual that says that D would lose his desire to drink from the glass if he learned that the glass contains petrol rather than gin. But, again, we can see that the mere appeal to this counterfactual cannot be sufficient, on a desire-based theory, to disclose what D has reason to do. We can see this by imagining an agent B who is like D in all relevant aspects (in particular, he shares D's aversion against drinking petrol) except that B's actual motivation includes a very strong aversion against acting on the basis of informed deliberation; B's motivational tendencies are dominated by his derision for the (as he sees it) petty, bourgeois obsession with risk calculation. Here it is true that B would not desire to drink from the glass if he found out that it contains petrol, but if we allow that this counterfactual discloses that B should not to drink from the glass there is nothing that prevents us from allowing that other counterfactuals disclose the

²³⁰ My objection to desire-based theories is related to but importantly different from the objection put forward, in different ways, by Hampton 1998 and Korsgaard 1997. Their objection is similar to mine in that they also claim the desire-based theorist has to make room for a kind of reason that is not based on desires. But for them this reason relates to 'the instrumental principle': they think that a desire-based theory must acknowledge that we have reason to take the means to satisfy our given desires that cannot be explained by appeal to desires. This objection presupposes that the instrumental principle as such has, or needs, a normative significance that is independent of the normative significance of the end toward which the relevant means are directed. This is far from obvious: the desire-based theorist might respond that that the *only* reasons for taking means to satisfy our desires are constituted *directly* by the desire in question (and, correspondingly, that we have *no* reasons for taking means to satisfy a desire that lacks normative authority). See my chapter IV, section 2 for further discussion.

normative practical reasons of agents even though those counterfactuals fail to relate to the actual motivational tendencies of agents. For instance, we might want to say that since it is true that D would not desire to drink alcohol if he were forced to attend a camp that educates people so that they despise the prospect of alcohol, D has reason not to drink alcohol, quite irrespective of the fact that D lacks any actual motivation to attend a camp of this sort.

Hence, the desire-based theorist must try to accommodate the intuition that D should not drink from the glass by appeal to the fact that D, unlike B, *has* a desire to engage in informed deliberation. But here the desire-based theorist runs into a dilemma. Suppose D does not check what is in the glass prior to drinking from it. On the face of it, this seems quite compatible with the idea that D's desire to act on the basis of informed deliberation *is* motivationally efficacious: being motivated to make informed choices cannot require omniscience or the need to check for outrageously improbable possibilities such as that of there being petrol in a glass on a cocktail table. But (this is the dilemma's first horn) in this case D's desire to drink from the glass *does* result from a process of informed deliberation and hence D *has* reason to drink from the glass. Against this, the desire-based theorist may hold that the notion of informed deliberation is, by definition, the notion of an idealized process that actually corrects, for instance, false factual beliefs; so we can say that D's fails to be motivated to engage in informed deliberation of the required, idealized kind. But if actually engaging in informed deliberation requires taking into account possibilities that seem bizarre and utterly improbable from the average deliberator's point of view in most circumstances ('can I really be certain that this glass doesn't contain petrol or acid?'), then it follows that average deliberators most of the time *lack* the desire to engage in this very special deliberative procedure. And this has obviously fatal consequences for the desire-based theorist's account to ground practical reasons in the appeal to informed deliberation.

But let us concede that D here does *not* act on the basis of informed deliberation, maybe because he does not act on the basis of deliberation at all: suppose D just rushes to the table to satisfy his thirst for alcohol without basing his choice on any deliberative effort. But (this is the dilemma's second horn) then, although D may be said to *have* a desire to engage in informed deliberation, this desire is not motivationally efficacious: his desire to drink outweighs his desire to step back and consider the situation in a more careful manner (whatever that would require). But then it seems like a desire-based theory implies that D *should not* deliberate in an informed manner. If the desire-based theorist responds that even though the desire to engage in (informed) deliberation did not actually motivate D on this occasion, it is nonetheless this desire rather than the motivationally efficacious desire to drink from the glass that discloses what D should do, then she commits herself to the notion that every agent has reason to engage in the activity of informed deliberation regardless of what her psychology happens to be like. But, as we saw, this commitment contradicts the desire-based theorist's commitment to the idea that practical reasons reduce to or are constituted by facts about an agent's psychology. Moreover, as we also saw, this commitment opens up the floodgates for discarding, quite generally, an agent's actual motivation as a guide to what she has reason to do and for appealing instead to counterfactuals about what she would be motivated to do in some scenario that, as a matter of actual psychological fact, she is not sufficiently motivated to be in.

The desire-based theorist might respond that what shows that D has no reason to drink from the glass is that *he would regret* drinking from the glass after having performed the action. But this would imply a commitment to the notion that every agent has reason not to act in ways

that she would, in hindsight, regret. This is, again, incompatible with the desire-based theorist's commitment to the notion that what an agent has reason to do depends on her psychological constitution, for an agent's psychological constitution may very well be such that she despises acting only in ways that she could or would not regret (consider again B and his anti-bourgeois outlook). Moreover, the condition that one may have a reason to do x only if one would not later come to regret doing x raises severe problems: what we might later come to regret is characteristically epistemically inaccessible to us, and it is not clear how late the 'later self' is supposed to be (for the desire-based theorist this is an especially problematic question because a later self may come to regret having done x precisely in virtue of a change between her later and her current psychological constitution; consider Parfit's russian nobleman).

The upshot of my argument thus far is that an appeal to counterfactual desires is not a legitimate way for desire-based theories to determine what an agent should do. But then we seem to be stuck, on such theories, with the appeal to the motivationally prevalent desire that actually moves the agent to act. This, however, is tantamount to the conclusion that the only criterion for singling out normatively significant (reason-giving) desires is the current strength (or motivational prevalence) of a desiderative state, which in turn implies the notion that every agent by default does what she has reason to do if she acts on the basis of her desires. And this is, as I argued above, tantamount to the abolition of the very concept of practical normativity.

Mark Schroeder, a recent defender of the desire-based theory, would hold that the objection I have expounded has no force against a sufficiently sophisticated version of this theory. He claims that the desire-based theorist is entitled to a notion of *correctness* of 'placing weight' on a reason that does not collapse into actual motivational tendencies: "...we...*know something* about what it is for a reason to be weighty. This doesn't have to do with whether people *actually* place weight on the reason, but it certainly has something to do with whether it is *correct* to place weight on it."²³¹ This raises the question of whether the desire-based theorist can give an analysis of *correctness* that does not reduce to either actual or counterfactual properties of desires (the appeal to counterfactual properties being ruled out by the argument thus far). On Schroeder's view the *weightiness* of a reason consists in the fact that it is *correct* for us to assign weight to this reason in our deliberation²³², and *correctness* is itself to be understood in terms of *reasons 'of the right kind'*.²³³ One feels like moving in a circle here, but my sense is that Schroeder's way out of this circle consists in the idea that the correctness of a reason (or its being 'the right kind of reason') is explained by reference to a specific form of activity:

The right kind of reasons involved in any activity are the ones that the people involved in that activity have, *because* they are engaged in that activity. So, for example, there are correct and incorrect moves to make in chess. The incorrect moves are ruled out, I think, by reasons to follow the rules of the game. Who has those reasons? Anyone who is playing chess. No one is playing chess, I think, unless she has some minimal level of desire to be following the rules of the game. And so anyone who is playing chess is guaranteed to have some reason not to, for example, castle out of check, because that is against the rules. People who are not playing chess do not have such reasons, but that does not matter.²³⁴

This sounds straightforward: if W plays chess, he has a reason to follow the rules that are

²³¹ Schroeder 2007, p. 101.

²³² Schroeder 2007, p. 129.

²³³ Schroeder 2007, pp. 133-4.

²³⁴ Schroeder 2007, p. 135.

constitutive of being engaged in this activity. But suppose that after playing chess for a while W suddenly feels an ardent desire d2 to make a move prohibited by the rules of chess. Suppose, say, that he feels attracted by the aesthetic configuration that would result from placing his king in the middle of the field. Should W follow the rules of chess? Schroeder says that he should since 'the right kind of reason' for W is dictated by the rules of the activity he is engaged in. But if all reasons are to be explained by desires, this reason can apply to W *only by virtue* of his desire for continuing to partake in the activity of playing chess. Now suppose that W's desire d2 (for rearranging the field in violation of the rules of chess) is much stronger than his desire d1 (for playing chess). I do not see how, within the confines of a desire-based theory, there is anything which accounts for the idea that W's reason for breaking the rules of chess is stronger than his reason for following the rules. But then the appeal to 'correctness' or 'the right kind of reasons' does nothing to replace the criterion of actual strength of given desires.

Now, the real question is how Schroeder's account is supposed to generalize from chess to cases where what an agent does is not obviously related to a form of activity that imposes a set of constitutive rules. His idea seems to be that we can appeal to *the activity of deciding what to do or of placing weight on reasons*.²³⁵ He does not show why or how this activity implies a set of rules that are analogous to the rules of chess, but he seems to think that it is somehow definitive of engaging in this activity that one does not place "inordinate weight on idiosyncratic reasons to take enormously costly means to fantastically frivolous ends", since that would be "a great way to undertake too many costs in order to accomplish the other things that [one] wants".²³⁶

What is the argument here? Consider again my above example where S strongly desires to cheat on his wife but also desires – less strongly, in the circumstances – to keep his marriage and to keep being employed by his wife's father. The question is on what basis a desire-based theories can normatively discard S's desire for cheating (D's desire for drinking from the glass, etc.). On one reading, Schroeder's point seems to be that acting on this desire would jeopardize other things that S desires. But it is *always* the case, for *any* agent with sufficient psychological complexity, that a given decision to do something prevents her from getting other things that she wants. Suppose I desire, to some degree, all of the following things: to see a movie, to go for dinner, and to have a peaceful evening at home. *Any* decision I make here will result in my not getting other things that I want to some degree, but this surely does not show that I should not go to the movies if I decide to do so. Schroeder might respond that there is reason to act in such a way that one can satisfy as many as possible of one's desires. But it is unclear how the normativity of this principle is to be explained – as it must be, on a desire-based theory – by appeal to properties of desires. Moreover, the principle is implausibly strong: deciding to assign priority to one desire among others seems generally acceptable (consider the artist who prioritizes his desire to produce art over his desire for marriage, a regular income, etc.).

Another response that is suggested by what Schroeder says in the above quotes is that no agent has reason to take costly means to *frivolous* ends that would jeopardize satisfying other ('non-frivolous') ends. But on what basis can we classify ends as ('fantastically') 'frivolous' (or reasons as 'idiosyncratic') within the confines of a desire-based theory? If, as seems plausible, the value of an end relates to what reasons there are for pursuing it, it follows, on a desire-based theory, that the value of an end for a given agent Z depends on what desires Z happens to have.

²³⁵ See Schroeder 2007, pp. 141-4. See also footnote 221 above.

²³⁶ Schroeder 2007, p. 143.

But then we are back to our familiar problem of how we can deem an end frivolous or whimsical for Z if Z as a matter of fact arduously desires to pursue this end.

Schroeder may, finally, say that since S engages in the activity of placing weight on reasons, his reason for cheating on his wife is 'the wrong kind of reason' given the fact that placing weight on this reason prevents S from getting many other things he wants. The problem is that Schroeder has done nothing to explain *why* it is supposed to be constitutive of the 'activity of placing weight on reasons' that one refrains from placing weight on a reason for acting in a way that would prevent one from getting many other things that one wants. There is no obvious analogy here between placing weight on reasons and playing chess. But let us concede, for the sake of the argument, that the constitutive link envisaged by Schroeder does exist. We would still face a problem that is analogous to the one we encountered in the above chess example. For even if placing weight on reasons to take very costly means to a given end (so that one cannot get other things one wants) *were* to violate a rule that is constitutive of the activity of placing weight on reasons, this could have normative weight for a given agent – within the confines of a desire-based theory – only to the extent that she desires to engage in that activity. And if she actually desires more strongly to break a rule imposed by this activity, we have to say that she should break this rule and hence should discontinue the activity of placing weight on reasons.²³⁷

Schroeder's appeal to a notion of 'correctness' of placing weight on reasons that does not collapse into actual or counterfactual strength of desires thus turns out to be a rather empty promise. Thus, I fail to see that Schroeder's sophisticated desire-based theory is more successful than the more straightforward account sketched by (say) Williams.

I want to consider a final naturalistic response to the worry I have raised, namely, Harry Frankfurt's appeal to higher-order identification. Frankfurt proposes to single out an agent's real ends by appeal to the agent's higher-order desires.²³⁸ Suppose, again, that S desires to cheat on his wife, *desires not to desire* cheating, and that S's first-order desire wins out. It is hard to see what gives S's second-order desire d2 normative authority over his first-order desire d1²³⁹: after all, d2 and d1 are on a par insofar as they are contingent psychological states, and if anything it would seem that d1 has a *greater* claim to reflecting what S 'really wants', in virtue of its motivational prevalence over the higher-order desire d2. The objection that it seems stipulative to grant authority to higher-order desires does not beg the question against desire-based theories: on such theories, a reason for S to act must be one that can motivate S to act, and given this supposition it is hard to see why desire d2 should 'normatively trump' desire d1 on a given occasion just because d2 is of a higher-order, if desire d2 cannot move S to act on this occasion.

Moreover, it does not seem true that second-order desires generally reflect what agents really want. Suppose V has fallen in love with a black man: she has strong, unreflective first-order desires to be held by him. But now suppose that she finds herself with an unreflective

²³⁷ Schroeder may respond that it is incoherent to say that an agent has reason to opt out of the game of placing weight on reasons. But there does not seem any incoherence here, on the assumption that what reasons an agent has are fixed by her desires. If she strongly desires things that imply or require opting out of the game of placing weight on reasons, then so be it; see Railton 1997 for a forceful defense of a similar point.

²³⁸ Frankfurt 1988b, p. 166: "Someone does what he really wants to do only when he acts in accordance with a pertinent higher-order volition." Lewis 2000 also confers authority to second-order desires insofar as he identifies 'valuing' with 'desire to desire'. Taylor 1985 appears to assign authority to second-order desires insofar as he seems to identify having such desires with evaluating our first-order desires.

²³⁹ See Quinn 1994, p. 192 and Watson 2003 for versions of the objection that the mere fact that a desire is of a higher order than another desire is irrelevant to the questions of its normative authority.

second-order desire not to feel attracted to black people, a residue of a subtly racist upbringing that she has suppressed.²⁴⁰ One might respond here that all this case shows is that we need to move one level further upward, i.e., that we must appeal to V's desire not to desire not to feel attracted to black people. But this raises two problems. First, we can imagine that V fails to form such a third-order desire, because she is not stable or critical or mature enough to engage in the kind of reflection on her upbringing that would be required for forming a desire not to desire not to feel attracted to black people. This does not at all show that she does not *really want* being held by her lover. Second, the appeal to third-order desires invokes the threat of a regress. If second-order desires aren't good enough for reflecting an agent's real ends, why should third-order desires be good enough?²⁴¹

Sometimes Frankfurt acknowledges the inadequacy of the appeal to higher-order desires. He explains that what is distinctive of the attitude of *caring* is not higher-order endorsement of a first-order desire, but that one takes an active stance toward the first-order desire (that one is "active in seeing to it that the desire is not abandoned or neglected"²⁴²). Now suppose I have spent months of work on building a hut in my garden. One day I come home from work, tired and stressed, and I hear that a storm is coming up. I know that if I do not cover the hut, the hut will be ruined; yet I end up watching mindless cartoons. Intuitively, this should be compatible with saying that I *care more about* finishing the hut than about watching cartoons. But it is hard to see how Frankfurt's analysis of *caring* supports this diagnosis. My present activity with regard to the desire for building the hut is *zero*; by contrast, I am active in seeing to it that my desire for mindless entertainment is not neglected: I grab the remote control, switch channels, etc. (Likewise, in the above example, S may be very active with regard to his desire for cheating: booking a hotel room, inventing subterfuges, etc.) One may respond that the point is that *in the past* I have been active in seeing to it that the desire for building the hut is not neglected. But the fact that one *was* active with respect to some desire *d* for an extended period of time can hardly show, of itself, that *d* *has* normative authority in the present, for it can surely be rational to abandon a past project in favour of a new desire.

²⁴⁰ For further counterexamples to and criticisms of the idea that valuing is a matter of having a second-order desire, see Scheffler 2010, pp. 18-9.

²⁴¹ Frankfurt 1988b, pp. 167 ff. tries to resolve a related worry by drawing an analogy between an arithmetic calculation and reflecting on one's first-order desires. The analogy is supposed to be that just as it would be unreasonable for me to continue with such a calculation ad infinitum unless I have some reason to suspect an error or to mistrust my calculation, so it would be unreasonable for me to continue reflecting on my desires unless I had special reason to do so. There are many problems here. First, if this response is good enough to warrant quitting reflection on one's desires after the formation of a second-order desire, it should also be good enough to preempt the need for such reflection and for the formation of a second-order desire in the first place. Second, the weakness of the analogy is that what warrants quitting the calculation is the appeal to an objective standard of correctness that I can reasonably take myself to have met; the problem of arbitrariness or circularity persists in the case of reflection on desires precisely because no such standard can be specified on the naturalistic account. Frankfurt says that "a person may be led to reflect on his own desires either because they conflict with each other or because a more general lack of confidence moves him to consider whether to be satisfied with his motives as they are." But this is rather elusive. Why should I care about a conflict in my set of desires? Is there something intrinsically good about consistency among desires? And on what basis can I settle the conflict if I cannot – as in the case of a conflict between beliefs – draw upon some kind of standard of correctness? Likewise, a 'general lack of confidence' about the adequacy my desires seems rational only qua worry about whether my desires are as they ought to be by appeal to some non-arbitrary, desire-independent justificatory standard.

²⁴² Frankfurt 1988b, p. 162.

The inadequacy of Frankfurt's proposals supports the overall point I have tried to establish in this section: namely, the point that the desire-based theorist seems stuck with the actual strength of desires as a criterion of normative authority. As I pointed out, this is a fatal consequence for a theory of normative reasons. It thus seems that the attempt of a reduction of normative practical reasons to desires (or to desires agents would have under certain non-actual conditions, or to desires toward which agents have been active, or to higher-order desires) fails.

3: Kant's incompatibilism about practical judgment

We saw, in section 1, that naturalists tend to agree with Kant about the incompatibility of the possibility of valid categorical or objective practical judgments with the assumption that all our practical thought is determined by contingent desires; they disagree with Kant in that they reject his claim that a naturalistic conception of practical reason implies the impossibility of valid practical judgments in general. Their rejection of this claim is, I suggested further (in section 2), driven by the following idea: the fact that our contingent desires control or determine our deliberative judgments about what we have reason to do does not tell against the validity or normative authority of those judgments, because the normative practical reasons represented by those judgments *themselves* depend on our contingent desires. In fact, I argued (at the beginning of section 2), desire-based theories thrive on the idea that both reasons and the deliberative recognition of reasons depend on the deliberator's contingent psychology: for unless the deliberative recognition of what counts as a reason for me is already (as Williams puts it) controlled by members of my 'subjective motivational set', one might suppose that the reason thus recognized is not dependent on my contingent psychology either.

The argument of section 2 shows that normative practical reasons, even those that *subjective* practical judgments purport to represent, are not metaphysically reducible to contingent desires. Hence, the psychological aspect of desire-based theories (their naturalistic account of, as Williams puts it, "what it is to come to believe" that one has reason to do x by appeal to the deliberator's desire-base) fails to correspond to their normative aspect (their metaphysical analysis of reasons in terms of desires). And this means that Kant's worries about the misfit between the naturalistic idea that our practical judgments are to be explained by appeal to our contingent desires and the idea that those judgments have normative authority over our contingent desires cannot be defused in the way envisaged by the desire-based theorist. On the naturalistic picture, all first-person deliberative judgments about what one ought to do are produced by some of the deliberator's given desires. We cannot (via the argument of section 2) single out any particular desire that the deliberator happens to possess as having normative authority over others. Hence, no deliberative judgment about what one has reason to do, which reflects the motivational grip that some particular desire currently happens to have on the deliberator, can claim to have authority over the various other desires that constitute the deliberator's psychology. (This point also applies, albeit in a different form, to third-person judgments about somebody else's reasons: such judgments, as we saw, derive – on the desire-based account – from the perception of someone else's desires; but via the argument of section 2, no mere perception of Tom's desires can yield a valid representation of what Tom should do.)

This abstract conclusion can be illustrated with the following example. Suppose that I judge that I should go to the dentist. On the naturalistic picture, this judgment is controlled by some of my desires – plausibly, by my desire to prevent the long-term, far-future pain that will

result from bad teeth. Now suppose that I am also afraid of dentists, and this aversion keeps me from actually going to the dentist. Intuitively, the desire to shun the dentist is irrational or capricious and does not constitute a conclusive reason to act for me. If the normative aspect of desire-based theories were adequate, such theories would allow us to say that the judgment that I should go to the dentist correctly represents how I should act and has normative authority over my capricious fear of dentists, because it is prompted by a desire that determines my 'real end'. But (via the argument of section 2) the idea that desires in and by themselves determine real ends or conclusive reasons to act fails. Hence, my judgment that I should go to the dentist is controlled by a desire that is normatively on par with the conflicting desire to shun the dentist, and this nullifies the judgment's claim to authority over the conflicting desire: the judgment fails as a valid representation of a normative practical reason. *If anything*, it looks as if my fear of dentists has, in virtue of its motivational prevalence, a greater claim to reflecting my 'real ends'.

One may wonder here why one cannot simply avoid this line of argument by rejecting the normative aspect of desire-based theories, i.e., by supposing that there *are* desire-independent facts of the matter about what there is reason to do that allow us to say that, in the above example, I ought to go to the dentist. But here I would insist, again, that this point cannot be endorsed without also rejecting the psychological, explanatory aspect of desire-based theories, namely, their account of how we come to endorse practical judgments. If it has to be conceded that what is causally responsible for our judgments about practical reasons are our contingent desires, then it is very hard to substantiate the idea that those judgments are responsive to desire-independent reasons. That is, if what prompts people to think that they ought to go to the dentist is their contingent desire to avoid future pain, then their deliberation and judgment is not responsive to, or expressive of, anything but their contingent psychology. Now, we are (dialectically) presupposing a (non-reductive) naturalistic conception of practical reasoning and judgment; and on such a view, I submit, the psychological aspect of desire-based theories really does give us the most plausible account of the phenomenon of practical deliberation that leads to judgments which are partly constituted by underlying dispositions to choose and act. It pictures practical reasoning as a function from empirically given desires to judgments about what we have reason to do that owe their practical nature (i.e., the fact that they are bound up with motivational tendencies, or dispositions to act) to the motivational input that gets practical deliberation going in the first place. On the naturalistic conception, there must be contingency all the way down, from deliberative input to our eventual evaluative thoughts and the resulting acts. My argument does not deny the explanatory coherence and potency of this picture, and its superiority over accounts that attempt to squeeze unanalyzed capacities for perceiving or intuiting the prescriptive force of necessary reasons and values into the naturalistic story. The problem is, rather, that if we picture practical deliberation in those naturalistic terms, then we cannot understand our practical deliberation as the recognition of desire-independent normative facts; and we cannot attribute normative authority over our contingent desires to the judgments that result from acts of reasoning that are controlled by contingent desires.

Now, one may further wonder how this argument could be consistent with Kant's own positive conception of the normativity of hypothetical practical judgments. Aren't such judgments, on his account, answerable to our contingent desires? Here I want to draw on the account of Kant's conception of the normativity of hypothetical practical judgments that I provided in chapter IV, section 2. Kant's account is indeed, in certain respects, congenial to

naturalists, for Kant accepts that as far as *non-moral* ends are concerned, what an agent has reason to do depends, in a deep sense, on her empirical nature: our desire for happiness is the only thing that counts (at least as far as the choice of *general* maxims and projects is concerned), and what gives us happiness wholly depends on our contingent psychology. Yet, as we saw, Kant thinks that even here practical reason has a saying that is irreducible to the influence of empirical desire (CPR, A 547-8/B 575-6): "Whether what is willed be an object of mere sensibility (the pleasant) or of pure reason (the good), reason will not give way to any ground which is empirically given." As I argued in detail in chapter IV, the crucial point here is that practical reason deems an act that is likely to secure an agent's happiness, i.e., her taking continuous pleasure in her own existence, (conditionally) *good*. This does not mean that practical reason here has genuine autonomy with regard to non-moral ends: it deems an act that secures an agent's happiness valuable because it finds that being pleased with our own existence is the *only* thing we (non-morally) care about and that we *always* care about it *equally at any given time*. But on this basis, it says that acts that allow us to get as much pleasure as possible, within the limits of moral permissibility, *ought to be done*; and it imposes a normative standard that is irreducible to the influence of empirical desires. What we desire at a given moment is not at all fine tuned to facts about what would or is likely to give us the greatest pleasure, and hence we do not by default desire what each and every one of us has non-moral reason to do. Non-moral practical reason, for Kant, is independent of *momentarily given* empirical desires. This independence, and the resulting standard of judging the advisability of satisfying some actually given empirical desires, underlies the normative force of hypothetical practical judgments on Kant's account.²⁴³

To sum up Kant's anti-naturalism about practical normativity: contingent psychological features are insufficient to ground any normative claims of practical reason (practical 'oughts'). But if naturalism is true, all of our practical reasoning and judging is determined by contingent psychological features. Hence, if naturalism is true, every verdict of practical reason is infiltrated and controlled by contingent psychological causes; and thus judgments of practical reason cannot claim normative authority over elements of an agent's contingent psychological constitution.

One final clarification is in order here: although I have tried to support Kant's incompatibilism or anti-naturalism about practical judgment by appeal to judgments representing subjective reasons (what Kant calls hypothetical imperatives), I do not think that Kant himself would resort to such an argument. He clearly thinks that the one and only important point here is that a naturalistic picture of the origin of practical reasoning and judgment cannot accommodate

²⁴³ How would Kant respond to one of the protagonists of the above argument, namely, B and his derision for prudential planning and 'informed choices'? Notice first that B's belief that living a daring life without regard to prudence is worthwhile seems to disregard the supreme authority of moral value, such as the notion that humanity (including the B's humanity) is an end in itself. Hence, Kant could point out that B suffers from an objective evaluative misconception, since B deems the value of her rational nature inferior to an obscure obsession with risk and danger. (It would not be question-begging for Kant to appeal to the categorical imperative at this point of the dialectic, for the desire-based theory that implies that there are no desire-independent practical necessities has already been shown to be problematic on independent grounds). If B somehow succeeds in combining his longing for risk and danger with respecting the moral law, Kant would say that this is only because her general attraction to danger (if not her concrete choices) are based on her taking pleasure in this kind of lifestyle, and then he can point out that it is irrational to engage in a lifestyle that is like to give one less of what one consistently wants (pleasure) and that her choices aim at a merely apparent good (the short-term kick she gets from the thrill of danger) as opposed to the one and only real conditional 'Wohl' (her own happiness). See chapter IV, section 2 for further discussion.

the possibility of a priori practical judgments or categorical imperatives, thereby rendering our moral practice illegitimate. I appealed to the case of hypothetical imperatives merely to break the stalemate we encountered at the end of section 1, that is, to show that the implications of a naturalistic picture of human reason are more serious than naturalists (especially those who think of our moral practice as built around mere fictions) are willing to acknowledge. Moreover, notice that Kant would insist that practical reason's role in prescribing non-moral ends is (i) very limited, given the fact that we do not know certain means toward our happiness and that cognizing these means is a task for theoretical reason (cf. Chapter IV, section 2) and (ii) in principle replaceable by the impact of desire or instinct: it is perfectly conceivable that our empirical nature is such that our given desires or instincts are infallibly attuned to what would make us happy²⁴⁴, in which case practical reason would have no non-moral employment.

4: Reflective distance and metaphysical independence

Rationalistic compatibilists such as Scanlon hold that as long as our reasoning and judgment is responsive to reasons, the fact that it is the causal upshot of antecedent natural conditions is completely irrelevant. I have argued previously (cf. section 1) that we cannot simply take for granted the idea that our acts of judgments are 'responsive to reasons' if we adopt a naturalistic worldview (specifically, a naturalistic conception of the causes of our acts of reasoning). I now want to confirm my skepticism about the attempt to combine an anti-naturalistic conception of practical normativity with the compatibilist stance of metaphysical neutrality toward the truth of a naturalistic conception of practical reason. Here I want to look at Christine Korsgaard's conception of the standpoint of practical reasoning. As we saw earlier (cf. chapter II, section 2), Korsgaard thinks that no metaphysical assumptions about the character of our capacities for reasoning and choice are needed to find substance in our normative commitments.²⁴⁵ One can endorse one's practical judgments as having genuine authority if one concedes that one makes those judgments only because one is psychologically 'wired' in a certain manner; one can even attribute normative authority to one's practical verdicts if one concedes that one's reasoning is controlled by an electronic device implanted in one's brain.²⁴⁶

I find it hard to see how this can be a consistent position. The question is how a judgment may claim to have an authority over psychological states that these states intrinsically lack if the judgment must be considered a product of those states. It seems as if the judgment could acquire this authority only from one of the factors under whose influence it is formed; and if those factors in themselves lack the required kind of authority it is not clear how they could produce a state which somewhat miraculously comes to express or represent it. Korsgaard says:

The reflective structure of human consciousness sets us a problem. Reflective distance from our impulses makes it both possible and necessary to decide which ones we will act on: it forces us to act for reasons.²⁴⁷ (...) Reflection gives us a kind of distance from our impulses which both forces us, and enables us, to make laws for ourselves, and it makes those laws normative.²⁴⁸

Korsgaard here appeals to a notion of reflective distance from our desires – a kind of distance

²⁴⁴ See GMS, 4: 395. I will return to this point in chapter VI, section 4.

²⁴⁵ Korsgaard 1996a, p. 183.

²⁴⁶ Korsgaard 1996a, pp. 162-3.

²⁴⁷ Korsgaard 1996b, p. 113.

²⁴⁸ Korsgaard 1996b, p. 129.

which gives our practical verdicts rational and normative *authority over* those desires. This surely is an accurate account of our self-conception as subjects of normative practical reasoning, an account that is very important to Kant's conception of human agency. The difference between Korsgaard's and Kant's position is that Kant does not accept that we are entitled to the notion of reflective distance from contingent aspects of our psychological constitution unless we are *metaphysically free from* the influence of those aspects in our acts of practical reflection and judgment. I agree with Kant that the notion of reflective distance from impulses is incompatible with the supposition that reflection is determined by such impulses. The notion of reflective distance surely does entail a kind of *independence* of the one who takes distance from the influence of those factors that she seeks distance from. And these are causal notions. It is unclear what is supposed to remain of the notion of reflective distance toward the contingent elements of one's psychological constitution if the judgments one purportedly makes from a position of reflective distance are formed under the causal influence of those very elements.

Korsgaard might respond that she is concerned with reflective distance from some particular natural state or desire *d1*, and all that this requires is that the reflection is independent of *d1*; it is irrelevant that the reflection is determined or controlled by natural states *d2*, *d3*, *d4* (where these states operate in the background rather than the foreground of reflection). But this response raises two problems. First, even if Korsgaard were entitled to the idea that adopting the reflective stance with respect to some desire *d1* suspends the causal influence that this desire has on our acts of reflection, she would have to accept that our reflection is (however covertly) controlled by other desires or elements of our contingent psychology (*d2*, *d3*, *d4*); and this is manifestly at odds with the claim that by adopting the reflective stance we can establish practical *laws* that have necessary validity, for this claim requires that our legislative acts enjoy independence from our contingent psychology *as such*. Second, given Korsgaard's firm agnosticism about the metaphysical and causal realities underlying processes of reflection, she is not even entitled to the idea that adopting the reflective stance with respect to *d1* suspends the influence of *d1* on our acts of reflection. If the (non-reductive) naturalistic conception of human reason is correct, then a process of deliberation about whether one ought to satisfy desire *d1* is determined by the deliberator's empirical psychology, as constituted by desires and inclinations of various kinds and orders, *including d1*. If the deliberative process leads to the verdict that *d1* is worth acting upon, this is simply proof of the motivational prevalence of *d1* or of some higher-order desiderative state based on *d1*; if the output of deliberation is that one ought not to act on *d1*, this reflects the influence of empirical desires other than *d1* (*d2*, *d3*, *d4*). If, as the naturalistic picture suggests, our empirically given desires are causally responsible for the very fact that we conceive of certain things rather than others as worth pursuing, then one cannot simply 'switch off' the causal influence of those desires by adopting 'the reflective stance'. Whatever we may believe, in the grip of our naïve self-awareness as reflective agents, about the sources of our reflective endorsement of practical judgments, the naturalistic picture entails that 'the laws we make for ourselves' are the causal upshot of our 'impulses', even if those impulses or desires do not operate in the foreground of our reflective consciousness.

Sometimes, Korsgaard suggests that the belief in our metaphysical freedom from natural causes that we adopt as reflective agents is dictated to us by practical considerations: we must adopt this belief so that we can act in accordance with our self-conception as autonomous agents. But I do not see how one can adopt this belief and remain agnostic about whether the

metaphysical realities underlying our acts of practical legislation are correctly described by the naturalistic picture as being determined by elements of our contingent empirical psychology.²⁴⁹

David Owens reads Korsgaard as a compatibilist and defends her "modest compatibilist notion of freedom" as follows:

Metaphysical libertarians will be unimpressed by this [Korsgaard's] picture of freedom. If our judgments as to what we should...do are themselves determined by factors beyond our control..., how, the libertarian will ask, can our capacity to make such judgments set us free? But the assumption lying behind this question, that true control must be total and unconditional, is unwarranted: control can emerge from a psychological background which we do not control. (...) For [Korsgaard], human beings have reflective control over their actions, have freedom of will, precisely because we can form a view of what we ought to do and implement our practical judgment in action. (Owens 2000, p. 11)

I agree that the libertarian (or, rather, incompatibilist) notion of control envisaged by Kant is not *per se* (independently of what one demands from a theory of freedom) preferable to the compatibilist notion of control envisaged by Owens. We can say, about an agent who exhibits this modest kind of control, that she is capable of figuring out whether she really wants to act on a given desire, in the light of her other given desires (the factors that, on the naturalistic picture, operate as the 'psychological background' of reflection), and that she is capable of doing or getting what she wants. But this is a very modest ambition for a conception of freedom. Kant's and Korsgaard's ambition is (as Owens rightly points out) to account for a conception of freedom that consists essentially in our capacity to form judgments about what we ought to do and to implement such judgments in our action (this is precisely Kant's notion of freedom as *autonomy*). And given this ambition, the concession that our judgments (and the choices based on these judgments) "emerge from a psychological background which we do not control" is rather problematic. If (as I have argued, drawing on Kant's ideas; and as seems implied by Korsgaard's notion of 'reflective distance') a judgment of what ought to be done purports to have a kind of authority over our empirical desires and (thereby) purports to represent (at least in the case of categorical judgments) a kind of practical necessity, I find it very hard to see how this ambition can be squared with the concession that the judgment is determined by (or 'emerges from') the empirical desires that happen to constitute our contingent psychological constitution.

In sum, the position that results from the combination of (i) an anti-naturalistic, rationalistic conception of practical normativity and (ii) a naturalistic metaphysics (or, at least, with a stance of metaphysical agnosticism that does not conflict with a naturalistic metaphysics) seems to me rather unstable. Endorsing Kant's ideal of autonomy comes at a metaphysical price: when Kant contrasts (GMS, 4: 448) a case in which reason can "regard itself as the author of its principles independent of foreign influences" with a case in which the subject of reasoning must "ascribe the determination of its judgment not to its own reason, but to an impulse" (GMS, 4: 448), the contrast is between a scenario in which the will qua practical reason (i.e., qua faculty of judgment) is metaphysically free from the influence of contingent psychological causes and a scenario in which our practical judgments are determined by such causes. It is only in the former

²⁴⁹ See Korsgaard 1996a, pp. 175-6 for the claim "that we must believe [that we can reflect and choose independently of all empirical conditions] in order to obey the categorical imperative". Unfortunately, in the same essay, she also claims (p. 162) that "the point is not that you must believe that you are free, but that you must choose as if you were free". I have previously (chapter II, section 3) discussed and rejected Korsgaard's attempt to reconstruct Kant's views on freedom as abjuring *a belief* in metaphysical freedom; below (in chapter VIII, section 6) I discuss the various strands and tensions in her published views on freedom in greater detail.

scenario that practical reason can be considered an autonomous source of necessary, a priori principles that apply to all rational beings.

In concluding, I want to notice two points. First, it must be emphasized that the freedom of practical judgment (from natural causes) that Kant envisages does not require alternative possibilities (cf. chapter III, section 3 and chapter IV, section 3). My argument in this chapter did not rely on the idea that it is in our power to 'judge otherwise'. Kant's incompatibilism about practical judgment requires not a freedom of alternative judgments but a freedom to establish objective necessities of practical reason: this freedom is incompatible with the causality of nature because one cannot establish objective practical necessities if one's practical reasoning is the causal upshot of contingent, rationally arbitrary natural states and events.

Second, we can see that nothing in Kant's incompatibilism about practical judgments hinges on his hedonism about non-moral ends. The incompatibilist argument I have expounded in this chapter does not depend on any specific features that our natural psychological constitution is supposed to have. Rather, it draws attention to the contingency and rational arbitrariness of the acts of practical reasoning and judging that (on a naturalistic picture) flow from this constitution. It is irrelevant for this incompatibilist argument whether our contingent moral outlook happens to be hedonistic or (say) intrinsically benevolent. An intrinsically benevolent moral outlook that was the causal upshot of contingent natural processes could not ground judgments that state objective practical truths or that have categorical practical authority. This shows that interpretations according to which Kant's incompatibilism stands and falls with his hedonistic conception of non-moral willing²⁵⁰ are rather shallow. It is the *contingency* of empirical desires that (on an exclusively naturalistic or deterministic worldview) determine our judgments and choices, *not* their intentional relation to *pleasure*, that drives Kant's worry.

This impression will be confirmed in the next chapter, where we discuss Kant's second incompatibilist worry about practical normativity. While the argument of this chapter was concerned with the legislative freedom that we exhibit qua authors of practical principles, the argument of the next chapter will be concerned with the 'executive freedom' we exhibit when we choose in accordance with or contrary to those principles. That is: the argument of the next chapter relates to the *formula* rather than the *content* of practical judgment, or to the *addressee* rather than the *author* of practical judgment, or to the act that is *governed* by practical norms (i.e., choice) rather than the act is responsible for *norm-giving*.

²⁵⁰ For such interpretations, see Irwine 1984, Williams 1986a, Foot 1972, and Wood 1984.

Chapter VI: Determinism and Objective Practical 'Ought'

In the preceding chapter, I examined what we might call the *legislative* dimension of Kant's incompatibilism about practical normativity: namely, his conviction that our practical judgments fail to have and express a normative content if they are caused by contingent psychological causes, such as empirical desires. In this chapter, I shall be concerned with what we might call the executive dimension of Kant's incompatibilism: namely, with his incompatibilist stance concerning the actions that comply or fail to comply with normative practical judgments (assuming that there are such judgments, and hence bracketing the worry examined in the preceding chapter). I will argue that Kant holds that an exclusively deterministic picture of human agency is incompatible with the idea that our acts are governed by practical laws.

1: Determinism, predictability, and imperfectly rational agency

I will begin by reviewing a common compatibilist view that one often finds associated with Kant. Proponents of this view rightly point out that it is essential to our agency that we act on the basis of deliberative choice. They suggest that determinism appears to threaten the deliberative stance via the *predictability* of our actions: adopting this stance requires that one takes oneself to have deliberative options, and this attitude would be precluded by a predictive belief that one *will* perform a particular act.²⁵¹ Those who understand the deterministic threat to agency in this vein assure us that the threat is in fact empty. Here they invoke the following line of argument. Suppose someone (myself, or a third party) tries to predict how I will act. Now, my learning of the prediction may well be causally relevant to how I decide to act (e.g., I might have counterpredictive motives). If so, the prediction is among the data that the predictor must use for inferring the prediction; and this means that the attempt to calculate the prediction presupposes, *per impossible*, that the prediction is known before it is calculated.²⁵² If, on the other hand, the prediction has no bearing on my action, this requires that the prediction does not interfere with my deliberative choice qua cause of my action, which in turn requires that the prediction is kept secret from me or that I set it aside in disbelief.²⁵³ So predictions of action are either impossible or irrelevant to deliberation.

This argument seeks to show that determinism is compatible with freedom, the relevant concept of freedom being *epistemic freedom*: for a deliberating agent, different actions are "possible for all [she] could possibly know."²⁵⁴ Proponents of this compatibilist strategy claim that it draws on a deep aspect of human agency, "the reflective structure of consciousness,"²⁵⁵ which allows us to treat a putative prediction about what we will do as a piece of information whose *practical relevance* it is up to us to decide. Thus, pace fatalism, determinism leaves intact

²⁵¹ See Bennett 1974, p. 213; Ryle 1949, p. 196; Korsgaard 1996b p. 95; Wallace 1994, pp. 3-4. For the view that if predictions of our actions were possible *in principle*, practical reasoning would be *in principle* dispensable as a way of figuring out what to do, see Bennett 1984, p. 111; Bok 1998, p. 79; and Korsgaard 1996a, p. 163.

²⁵² See Bok 1998, p. 81: if the prediction were itself a potential cause of action, the predictor would have to "factor the result of [her] calculation into [her] calculations in order to arrive at a result, and [she] would have to know what that result was before arriving at it in order to do so." See Korsgaard 1996a, p. 163 and 1996b, p. 96, MacKay 1960, Popper 1950, Ryle 1949, pp. 196-7, and Scriven 1968 for variants of this argument.

²⁵³ Arguably, I cannot deliberate about *whether to do x* and choose to do x on the basis of my deliberation if I already believe that I will do x. See Bennett 1974, p. 217; D. Perry 1965, p. 239; Wallace 1994, p. 3.

²⁵⁴ Bok 1998, p. 120. See also MacKay 1960, pp. 31-32; Moore 1912, p. 94; Wallace 1994, pp. 3-4.

²⁵⁵ Korsgaard, 1996b, pp. 96-7. Compare Bok 1998, p. 108, pp. 120-1 and MacKay 1960, p. 40.

deliberative choice as an indispensable condition of agency that prediction cannot preempt. This is meant to show that metaphysical issues are strictly irrelevant to our practical self-conception.

Is epistemic indeterminacy enough to account for our sense of freedom of action? Imagine a computer, sensitive to linguistic input, that has been programmed to act counterpredictively: whenever it calculates or is told that it will do *x* at time *t*₂, this causes it to do something other than *x* at *t*₂. This machine cannot calculate its future activity²⁵⁶, and a third party can predict the machine's output only if she does not announce her prediction to it. The causal output of this computer is as *epistemically open* as our actions are. But is the computer *free*? Admitting its freedom would involve a radical clash with our self-conception: we think that our freedom is not shared by mechanical devices. So compatibilists should deny that the computer is free since it lacks the power of deliberative choice. This, however, shows that the immunity of forms of causal output against preemption by prediction is independent of any deep aspect of human agency. This immunity rests on a feature that may be instantiated by the 'decision procedures' of mere machines. This point does not yet show that (or why) our self-image as human agents requires a notion of freedom that is thicker than the epistemic concept of freedom; but it should at least make us wary of the idea that the epistemic indeterminacy of our actions is sufficient to account for the sense of freedom that shapes our experience of agency.

This brings me to Kant's position. It is often suggested that for Kant, our freedom rests on the fact that our actions are epistemically open to us²⁵⁷, but this idea is hard to sustain. Notice first that Kant considers human actions strictly unpredictable because he thinks that we cannot cognize psychological laws and causes.²⁵⁸ Surely, if Kant *did* think that determinism threatens our agency via the predictability of our actions, he would dismiss this threat as empty by appealing to our inability to predict these actions. But his defense of freedom *never* proceeds on this basis. Rather, in his discussion of freedom Kant abstracts from our epistemic limitations and imagines a scenario in which our actions *are* predictable.²⁵⁹ This strongly suggests that his worry about our agency in a deterministic world is not at all driven by concerns about predictability. Likewise, if Kant *did* think that we can ground our freedom in the fact that our choices are necessary conditions of action that prediction cannot preempt, he would accept that our actions are free if they follow from our choices, *even if* these choices result solely from deterministic causes. But for Kant, such compatibilism would give us only the "freedom of a turnspit".²⁶⁰

²⁵⁶ The computer might never finish a calculation of what it will do because it gets stuck in an infinite loop; or it might be programmed in such a way that it does not factor its counterpredictiveness into its calculations.

²⁵⁷ See Bennett 1984, p. 111; Hill 1992, p. 136 ff.; Korsgaard 1996a, pp. 162-3; Wallace 1994, pp. 3-4. See Korsgaard 1996a and Matthews 1969 for the related view that Kant's solution to the problem of determinism lies exclusively in an appeal to a metaphysically neutral practical standpoint (I provide a thorough critique of this line of interpretation in chapter II). A very different version of the idea that epistemic considerations are crucial to Kant, which leads to *libertarianism* rather than compatibilism, is suggested by Hogan 2009. I agree with Hogan that Kant believes that the noumenal grounds of our actions are unpredictable because they are undetermined, but I do not see that this point plays a central role in Kant's theory.

²⁵⁸ For instance, Kant thinks that an agent's awareness of her own or somebody else's intentions to study her volition prevents the experimental conditions for cognizing the (existing) psychological laws that govern the empirical or phenomenal grounds (i.e., psychological motive causes) of our actions (Man, 4: 471). See my chapter II, section 2 for further discussion of these issues.

²⁵⁹ CPR, A550/B578: "...if we could exhaustively investigate all the appearances of men's wills, there would not be found a single human action which we could not predict with certainty and recognize as proceeding necessarily from its antecedent conditions..." See also KpV, 5: 99.

²⁶⁰ See KpV, 5: 97. See my chapter II (especially section 3) for an extensive argument to the effect that Kant's

So Kant's claim that we must "act under the idea of freedom" (GMS, 4: 448) cannot be interpreted in terms of mere epistemic freedom. But what, then, *is* his point here? Why *should* it matter to us whether we are causally determined?²⁶¹ I claim that Kant believes that if our actions were solely governed by natural laws, they would be deprived of their *objective normative* dimension, because 'ought' judgments that represent laws of practical rationality would lose their point and meaning.²⁶² By an *objective* norm or *law* of practical reason, I mean a norm that gives necessary reasons for acting that apply to every agent.

What is Kant's argument here? Let determinism be the view that every event that has causal efficacy is exclusively governed by deterministic natural laws and so necessitates its effect(s).²⁶³ Now it might be the case that our deliberative reflection on a norm cannot, as a matter of natural necessity, have any causal efficacy. If so, then either we are necessitated to comply with the norm or we are necessitated to violate the norm, irrespectively of our reflection on that norm. If, on the other hand, reflection on a norm *has* causal efficacy, such reflection necessitates either compliance with or a violation of the norm. In neither of those scenarios is it possible, in an incompatibilist sense that tracks the absence of causal necessity, for us to go either way with respect to (i.e., accept or reject) a norm. For the present purposes, assume that if determinism is true, these are the only possible scenarios. It follows that if determinism is true, it is never possible (in an incompatibilist sense) for us to go either way with regard to a norm. (I shall consider compatibilist interpretations of 'can' in section 2.1.)

For Kant, this implication of determinism conflicts with the idea that our rationality is Janus-faced: we can act on the basis of our appreciation of the force of normative laws, but this ability has an ontological *shadow or privation*, namely, our propensity to violate the demands of reason.²⁶⁴ We are rational, but *imperfectly* so; our recognition of a law of practical reason leaves us with the option of non-compliance. By contrast, a *perfectly* rational being, such as God, cannot but comply with practical laws. The difference between the way in which perfectly and imperfectly rational agents are influenced by their recognition of normative laws comes to light in the fact that only the latter apprehend such laws in the form of 'oughts'. The point of 'ought' is to represent what it would be right to do to agents who may choose the right *or* the wrong thing.²⁶⁵ If determinism is true, human agents are, whenever they reflect on a normative law in

insistence that a naturalist (qua transcendental realist) is not entitled to his theory of freedom precludes a compatibilist or a metaphysically deflationary reading of his theory of freedom.

²⁶¹ Bennett 1974, 1984 suggests that the appeal to putative predictability is the *only* way to make sense of Kant's point. I agree that there is an interpretive burden here. Commentators usually fall back on the issues of moral responsibility or moral worth (Bojanowski 2006; Irwin 1984; Wood 1984), but it is not clear in what way these issues could threaten our agency on the whole. Allison 1990, pp. 38-41 traces Kant's worry to the idea that a deliberating agent must consider herself as having a kind of control over her choices, but he does not explain what this control amounts to or why it requires the absence of causal necessitation.

²⁶² Pace Pereboom 2005, I take Kant's point to be that a universal determinism would render objective 'oughts' (categorical imperatives) *meaningless or impossible* rather than just *false*. CPR, A 547/B 575: "When we have the course of nature alone in view, 'ought' has no meaning whatsoever. It is...absurd to ask what ought to happen in the natural world." MS, 6: 222: "The ground of the possibility of categorical imperatives lies in the fact that they refer to no determination of the activity of the will... but solely to its freedom."

²⁶³ In Kant's terms, let us proceed as if "all causality were mere...nature" and hence that "every event [were] determined by another in time in accordance with necessary laws" (CPR, A 534/B 562).

²⁶⁴ See MS, 6: 227 for the idea that the 'ability' to violate laws of reason has the ontological status of a privation.

²⁶⁵ GMS, 4: 413-4: "All imperatives are expressed by the word ought, and thereby indicate the relation of an objective law of reason to a will, which from its subjective constitution is not necessarily determined by it...."

their deliberative course, deprived either of the option to comply with or the option to violate the law. Hence, if determinism is true, objective 'oughts' are pointless. Since the objective normative dimension of our agency depends on the applicability of meaningful objective 'oughts' to our actions, determinism implies that our actions are not governed by meaningful objective 'oughts'.

We can also see Kant's point by examining the way our actions are caused in the deterministic scenarios described above. If the causal nexus is such that reflection on norms *cannot* influence action, the action obviously lacks normative significance. If reflecting on a norm necessitates a violation of the norm, this influence of deliberation on action cannot be considered rational: a rational influence cannot necessitate the violation of a law of practical reason. This leaves us with the case where apprehending a normative law necessitates compliance. Deliberation here may comprise separate events (a weighing of options), but attendance to the norm is what brings about the act. Now, a *perfectly* rational being's recognition of the rational force of normative laws leaves it with no option but to do what the law says must be done. But when an imperfectly rational deliberator recognizes a normative necessity, any *rational* as opposed to merely causal influence would leave her with the option to violate the norm representing this normative necessity.²⁶⁶ So, if the deliberator *is* deprived of this option, the influence that leads her to act is the merely causal influence of natural, nomological necessity: she is determined to act by non-rational factors and causes. Hence, if determinism is true, human actions are in every case produced by a non-rational, merely causal influence, and hence they lack the normative dimension that meaningful objective 'oughts' would provide for them.

One initial objection to this argument is that it assumes that in a deterministic world causally efficacious events must individually necessitate their effects. Couldn't an effect be caused by a cluster of events that are individually necessary but only jointly sufficient for its occurrence? This would give us a scenario in which reflection on a norm brings one closer to compliance without being sufficient for and hence without necessitating compliance.²⁶⁷ But this idea raises a grave problem of its own. If reflecting on a norm is insufficient for an act of

They say that something would be good to do or to forbear, but they say it to a will which does not always do a thing because it is conceived to be good to do it. (...) A perfectly good will would therefore be equally subject to objective laws... but could not be conceived as obliged thereby to act lawfully, because of itself from its subjective constitution it can only be determined by the conception of good. Therefore no imperatives hold for... a holy will; ought is here out of place, because the volition is already of itself necessarily in unison with the law. Imperatives are only formulae expressing the relation of objective laws of volition in general to the subjective imperfection of the will of this or that rational being, for example, of the human will." Compare MS, 6: 222.

²⁶⁶ Kant says that imperfectly rational agents can be *genötigt* ('necessitated') by objective reasons (GMS, 4: 412). He sometimes simply equates such 'Nötigung' with obligation (KpV, 5: 81). It is a very special notion: it differs from *natural* necessitation, and Kant *also* deems it inapplicable to God. Our 'necessitation' by laws of reason is such that, as Kant puts it (GMS, 4: 412), our "will from its nature does not of necessity follow" these laws, so that our actions are "subjectively contingent." In order for an imperfect will to stand in the relation of *Nötigung* to an objective law of reason, the following conditions must be satisfied: (i) the law *suffices* to move ('determines') the will through a rational influence experienced as *constraint* (KpV, 5: 92); (ii) the will has the genuine option of defying the law. Determinism negates (ii); (i) would be negated by the assumption that causes are *indeterministic*. (I shall elaborate a bit on this last point in what follows; see footnote 268 below.)

²⁶⁷ Another type of case I have left out of account is *overdetermination*. An overdetermined act of compliance with a norm in a deterministic world is necessitated by an apprehension of an ought and it would also necessarily have occurred if the ought had not been apprehended. Here the agent does not have the freedom to respond in either way to the norm, and the act of compliance would have occurred regardless of her reflection on the norm. Each of these points seems sufficient to show that the 'ought' is pointless.

compliance, the act is produced partly and essentially by events that are independent of the agent's rational self-governance; in this case the act is not fully attributable to her and *a fortiori* lacks a normative dimension. One might insist that in this case the act has some degree of normative significance which corresponds to the degree to which the agent's reflection on norms contributes to her act. But such talk about degrees of rational control is obscure and has no repercussion in our ordinary thought. It cannot provide a substitute for our conviction that appreciating the force of objective reasons can be a sufficient ground of our agency.²⁶⁸

One might respond that the appeal to a variety of events as joint causes of actions misses the point: the point should, rather, be that an act of compliance necessarily occurs under the hypothesis that the agent has certain causal dispositions whose presence (and realization) in the circumstances necessitates compliance with the relevant norm. This suggestion avoids the worry that *events* other than rational self-control are needed for compliance (or non-compliance). But I do not see how it even addresses the worry raised by Kant's argument. If an agent's act of compliance with a norm occurs necessarily (in the circumstances) as a result of her reflection on reasons, *given her inner constitution* (i.e., her psychological dispositions), this means that her reflection on reasons necessitates compliance with the norm. Now, Kant's point is precisely that an imperfectly rational agent cannot determine herself to act as she ought to in such a way that her appreciation of the normative force of reasons leaves her with no option but to act on those reasons. Thus, if the internal dispositions that constitute the agent's psychological nature (her 'empirical character', as Kant would put it) leave her with no genuine option but compliance, her compliance cannot be due to the rational influence exerted by her appreciation of conclusive reasons: rather, it is the upshot of the influence of psychological or neurological factors that fall under natural laws and that belong to the space of causes rather than to the space of reasons.

But Kant's argument does face serious challenges. I think that there are three main objections to Kant's incompatibilist line of thought, which I shall consider in what follows.

2.1: Compatibilist interpretations of 'can'

I first want to consider the compatibilist response that having the double option to comply with or to violate practical laws does not require us to be free from causal necessitation.

The compatibilism we encountered in section 1 and which invokes an *epistemic* sense of possibility fails to address Kant's argument. This argument concerns the causal conditions required for an objective normative dimension of human agency, or the ontological abilities we must possess to qualify as addressees of objective 'oughts'. The fact that we typically can comply with or violate objective norms *for all we can know* is irrelevant to the worry that we never *in fact* have this double option (i.e., the possibility to go either way) and that our acts are therefore never governed by the practical 'oughts' which presuppose this double option.

We thus need to consider compatibilist views that address not our epistemic situation but our real powers. Here one influential suggestion is to analyze 'can' in conditional terms: to say that R *can* do x is to say that R would do x if she chose to do x. Kant dismisses this idea as a "wretched subterfuge" (KpV, 5: 96). Less polemically, the problem with the conditional analysis is that it provokes the question of whether R *can* satisfy the mentioned condition, i.e., whether R

²⁶⁸ This relates to the point I made about the implications of the idea that our actions are exclusively subject to indeterministic natural causality in footnote 266 above. This implies that our reflection on reasons can only ever increase the likelihood of our acceptance of or rejection of practical norms; it can never suffice for an act of compliance or non-compliance, so that our actions are never fully within our rational control.

can choose to do x. If R cannot choose differently, it seems pointless to say that R would act differently if she chose to do so. In response, the compatibilist can offer a further condition under which R can choose²⁶⁹, but this gives rise to a fatal regress: any condition that is offered here calls again for the original question, namely, 'can this condition be satisfied?'.²⁷⁰

This difficulty motivates a compatibilist approach that refrains from analyzing the notion of 'can' and instead focuses on a *general, unanalyzed* notion of a capacity. It emphasizes that human agents (unlike, say, donkeys) have a general ability to respond to objective norms that is not threatened by determinism.²⁷¹ Proponents of this view could add that we also have a general 'ability' to violate norms.²⁷² Now, the problem here is that such general abilities could not have significance for us apart from the *possibility to exercise them* in concrete acts. And if we seek to delineate an agent's genuine options on a concrete occasion, the appeal to general abilities is not sufficient: for instance, the general capacity to run can be ascribed to a healthy man (as opposed to someone who lost his legs) even if he is chained and thus *cannot* exercise it.²⁷³

The compatibilist can respond that the attribution of a general capacity *is* sufficient to delineate what an agent is free to do unless there are preventive conditions, such as chains, that block the exercise of this capacity; and determinism does not imply that such conditions universally obtain.²⁷⁴ To assess whether this is an adequate response, we must consider what makes chains a preventive condition. One plausible suggestion is this: the claim that R cannot exercise his ability to run at time t1 if he is chained involves a reference to what the laws of nature (dis)allow; given factors such as the mass of the chains, R cannot (as a matter of nomological necessity) run at t1. Now, *if* this account of preventive conditions is on the right track, determinism *does* seem to imply the proliferation of such conditions. Suppose R chooses to lie at time t2. If determinism is true, all events, and hence the exercise of all of R's general capacities including his ability to tell the truth, are governed by deterministic natural laws. Thus, if R does not exercise this ability at t2, it is (nomologically) impossible, given neurological factors such as R's brain state, for R to exercise this ability at t2 and to choose not to lie.

It may be replied that our *ordinary concept of an obstacle* applies to chains but not to brain states. But the fact that we tend to assume that neurological conditions do not impose insuperable constraints on deliberation and choice hardly shows that this assumption is legitimate in a deterministic world. It may be based on a pre-theoretical libertarian intuition or on our near to total lack of knowledge of neurological laws. Our ordinary usage of the concept of an obstacle

²⁶⁹ Moore 1912, p. 94 proposes to re-apply the conditional analysis to 'can choose'; this would give us the (rather artificial) notion 'R can choose to do x if: R would choose to do x if she chose to choose to do x'.

²⁷⁰ See Wolf 1980, p. 154: "...this debate might be carried on indefinitely with a proliferation of conditionals and a proliferation of objections. But if an agent is determined, no conditions one suggests will be conditions that an agent could have satisfied." See also Chisholm 2003, pp. 28-30. As we saw (cf. p. 114), a number of philosophers apply the epistemic sense of 'can' to 'can choose' to cut off the regress, but (as we also saw) epistemic possibility is not enough. Davidson 1973 holds that the conditional analysis should focus on what one *can desire* (rather than on what one can choose), desires being (according to his doctrine of 'anomalous monism') undetermined; but (cf. Wallace 1994, p. 193) we do not attribute free actions to mere desires.

²⁷¹ See Frankfurt 2003a; Reath 2006, pp. 255-6; Scanlon 2003; Watson 2003; Wolf 1990, 2003.

²⁷² Wallace 2006 emphasizes the significance of a 'counter-normative capacity.' Wolf 1980 and 1990, by contrast, argues that we need only posit an ability to choose *correctly*. I consider her view below in footnote 286.

²⁷³ Wolf admits (2003, p. 387) that "according to our commonsense understandings, having [an] ability is one thing and exercising it is another." See also D. Locke 1973-4, p. 179, and van Inwagen 1983, p. 13.

²⁷⁴ See P. Strawson 2003, p. 87 and Wolf 1990, pp. 111-2.

or a preventing condition cannot be cited in defense of compatibilism because it *presupposes* that there is a categorical difference between cases in which we are constrained not to exercise some of our general capacities (e.g., when we are chained) and cases where we are free to exercise all or most of our general capacities. It is hard to see what may ground this difference if agents are on every occasion causally determined to exercise just one among their general capacities and if the exercise of all their general capacities is always governed by deterministic laws.

The argument just given draws on the assumption that, *if determinism is true*, the sense in which we can or cannot exercise our general abilities is fixed by what the laws of nature allow. It might be objected that this appeal to nomological modalities, in defining what our options are, begs the question against compatibilism. But then the compatibilist owes us an account of the notion of the possibility to exercise general abilities that does *not* require an appeal to natural laws. Here it might be said that we can use the conditional analysis to specify why R, if chained, cannot exercise his ability to run: R cannot run because he would not run if he chose to do so.

But this suggestion raises two problems. First, the claim that R, if chained, would not run if he chose to do so *is* sensitive to our understanding of what is physically possible given the way in which the laws of nature operate; the claim would be false in a universe where the laws of physics differ greatly from those that govern events in our actual world. Hence, the conditional analysis does not replace but rather relies on an (implicit) appeal to physical possibility. The second problem is this: if our reason for deeming it impossible for R to run if he is chained is that he would not run if he chose to do so, then it follows that our reason for deeming R free to run in the absence of chains is that he would run if he chose to do so. But then it is the conditional analysis rather than the notion of a general capacity that is doing the work here; and, as we saw, the compatibilist appeal to general unanalyzed capacities was motivated in the first place by the justified worry that the conditional analysis is ultimately deficient (i.e., regressive).

Thus, neither the fact that we would do something different if we made certain choices nor the fact that we have the general ability to make certain choices seems by itself sufficient to show that we have the genuine option to make those choices *under the assumption that* the exercise of our powers of choice is exclusively governed by natural laws. There are, of course, further compatibilist proposals that I cannot consider here in detail.²⁷⁵

To sum up: the argument (from the possibility of objective 'ought') of section 1 provides a (distinctively Kantian) foundation for the intuition that freedom requires alternative possibilities.²⁷⁶ The argument of this section supports the view that compatibilist accounts of alternative possibilities are not satisfying.

²⁷⁵ There is, for instance, 'Lewis-style compatibilism.' Lewis 2003 suggests that the sense in which we *can* exercise general powers that we do not in fact exercise is that in which we can do something that *entails* but *does not itself cause* a violation of the laws of nature. But this is a mere technical distinction whose relevance for our deliberation is obscure (Beebe 2003 argues persuasively that Lewis' distinction tracks *no* relevant difference). Moreover, one wonders whether the attribution of powers to do something that entails a breach of natural laws is really as metaphysically innocuous as Lewis makes it sound. How far do such powers extend? Could the chained man do something that entails a 'local miracle' which allows him to break his chains? Lewis would respond that a possible world in which someone makes a non-actual choice is 'closer to the actual world' than one in which someone breaks out of chains. But this suggests that our powers of choice are, as a matter of *de re* modality, not as tightly constrained by the laws of nature as our movements of limbs; and *libertarians* may find this appeal to a kind of metaphysical possibility that is unconstrained by physical laws congenial.

²⁷⁶ Dennett 1984, Frankfurt 2003b and Wallace 1994 reject the significance of alternative possibilities; however, they do so in the context of discussing responsibility rather than forward-looking agency.

2.2: Two different conceptions of rational imperfection

I now want to consider the worry that Kant's argument depends on an idiosyncratically strong construal of the idea of rational imperfection. For Kant, this idea implies that human agents can violate objective norms on each and every occasion. But it might be held that we can plausibly interpret this idea in a weaker sense that is compatible with determinism, by invoking two features. First, even if human agents are at some one time (on a particular occasion) necessitated to comply with normative laws, we may account for their imperfection by considering their agency *across time*, i.e., by picking out past instances of actual non-compliance. Second, even if our reflection on a law necessitates us to comply with the law, the necessity is *hypothetical*: compliance with a norm that is caused by deterministic processes depends on the contingent history of the world and on the character of the laws of nature. It is always true that we *conceivably might fail* to comply. By contrast, God's compliance with practical laws does not depend on contingent factors and His non-compliance is inconceivable. So it seems that we can honor the fact that 'oughts' can meaningfully be addressed only to imperfectly rational agents without requiring that such agents *always* have the option to reject objective norms.

I shall refer to the conception of rational imperfection inherent in this objection as Weak Imperfection.²⁷⁷ The defender of a more radical conception of our rational imperfection, who holds that we always possess the option to violate laws of reason, must show that Weak Imperfection is unsatisfactory. To show this, I shall adduce two different considerations.

Weak Imperfection allows us to locate a causally determined agent's imperfect rationality in the relation between her activity and norms across time or in the fact that her non-compliance is conceivable. But it downplays the impact that our rational imperfection makes *in a concrete deliberative context* where our appreciation of objective reasons leads us to choose correctly. It allows that if an imperfectly rational deliberator complies with a norm because of her reflection on the norm, her considering the norm removes her option to choose the wrong thing. This, however, negates the idea that even in those cases where our deliberation leads us to choose correctly, the conditions that render our rationality imperfect (proneness to temptation, self-deception about our motives, etc.) inveigh against the rational influence exerted by our recognition of normative laws. Weak Imperfection implies that in such contexts, the conditions that seek to divert us from choosing correctly have no claim on us: they are rendered powerless by our recognition of normative laws. One might say that the influence of these conditions is not negated because the right choice occurs as the result of a process of deliberative competition between various motives. But the idea that the factors that account for our rational imperfection genuinely compete against our power to choose for the right reasons is inextricably tied to the idea that these factors have a chance of success, i.e., *may sway us* to choose the wrong thing. If it is impossible for us to choose the wrong thing, there is no such chance of success. Thus, notions such as struggle, temptation or weakness of will are rendered inapplicable, for they are tied to our awareness of a possibility to choose contrary to our apprehension of the demands of reason.

One might respond that Weak Imperfection leaves room for such self-awareness, since it

²⁷⁷ Weak Imperfection should be congenial to Wolf 1980, 1990 or to Pettit/Smith 2003. It might also be attributed to Lavin 2004. He claims (p. 441, 443) that it is a trivial, conceptual fact about the term 'imperative' that being subject to imperatives requires the possibility of violation. Lavin never quite spells out what kind or concept of 'possibility' is employed in this allegedly 'trivial' idea, but he surely (cf. *ibid.*, p. 448) favors a very modest conception. Weak Imperfection also seems to figure in McDowell, who appeals (1998b, p. 105) to a "potential gap" between ideal and actual motivation; this vague, modest idea fits perfectly into a deterministic framework.

leaves room *for the experience of* temptation or struggle: a deterministic causal process that necessitates an agent to comply with a norm may involve, as parts of the causal chain, events that cause such experience. While this is correct, there would be something illusory about the experience of struggle or temptation in a case where one is necessitated to do the right thing: it is constitutive of genuine temptation and struggle that what one is tempted by, or what one struggles against, can sway one to act; and thus it is constitutive of the experience of temptation that one takes oneself to be aware of a real (not merely epistemic) possibility that what one is tempted by can sway one to act. For instance, if Tom feels drawn to cheat on his beloved wife, his experience of temptation is constituted by a lively awareness that he may give in to his carnal desires – i.e., that it is actually possible for this to happen, not merely for all that he can know.

Now, why couldn't the defender of Weak Imperfection be happy to account simply for the *illusory* experience of temptation (in cases where the agent is necessitated to choose correctly)? Recall that Weak Imperfection was offered in an attempt to show that doing justice to the idea of rational imperfection does not require Kant's radical interpretation of this idea. This ambition cannot be reconciled with the concession that Weak Imperfection requires us to regard characteristic aspects of our practical experience as mere illusions.

There is a second consideration that tells against Weak Imperfection. We saw that it tries to locate our imperfect rationality in two features: first, in non-compliance across time; second, in the fact that compliance is always contingent on the history of the world and the character of the laws of nature. Now, the first of these features cannot be *definitive* of our imperfection: the extent to which we are non-compliant across time is itself contingent whereas our rational imperfection is a necessary aspect of the human condition. It is only the second feature that allows Weak Imperfection to specify an idea of imperfection that applies to *any* human agent regardless of her contingent place in the causal nexus. Now, imagine Sara who, due to her place in the causal nexus, continuously and without exception does the right thing and is necessitated to do so by her reflection on the relevant norms. Sara differs from The Holy Will in that her immunity against temptation depends on contingent factors that might conceivably not have obtained. But how could an appeal to a hypothetical scenario in which she *would be able* to act contrary to reason give substance to the notion that she *actually is* imperfectly rational?²⁷⁸ (I am presupposing here, via the argument of section 2.1, that compatibilist interpretations of what an agent can do – regardless of her being causally necessitated – are unsatisfying.) The conditions that grant her constant immunity against acting wrongly are satisfied and their satisfaction is irreversible, the history of the universe and the laws of nature being the way they are. Yet, Sara counts as imperfectly rational in the *one* sense that is, according to Weak Imperfection, definitive of *what it is to be* imperfectly rational for human agents as such. If we do not recognize in Sara what we think is essential to our sense of our own imperfection, this tells against Weak Imperfection. It may be responded that we *can* recognize our imperfection in Sara – we just happen to be less fortunately situated in the causal nexus. But this suggests that the actual

²⁷⁸ One might suggest that we can give substance to this notion by conjecturing, 'If Sara had been subjected to a certain temptation at t1 then her attendance to a norm would not have caused compliance.' But we can picture Sara as being psychologically constituted in a way that *no* such counterfactual comes out true. Of course, we can appeal to a possible world in which her psychological constitution is different from the one she actually exhibits and includes susceptibility to temptation. But this just reiterates that it is *conceivable* that Sara might act wrongly; it does not allow us to locate Sara's imperfect rationality in the actual world and with regard to the way in which her actual powers constrain her choices.

possibility to act wrongly is an accidental feature of our human condition, whereas the conceivability of acting wrongly yields the essence of our self-conception as imperfectly rational; and this reverses the real order of ideas. It is because we have an acute awareness of the actual possibility that we may act wrongly that the conceivability of acting wrongly has such a strong grip on us; and there are many things that are merely conceivable and *therefore* have no bearing on our self-conception.

I do not mean to imply that appeals to what is conceivable or what may occur in a possible world are generally pointless; that is, I do not want to rule out that such appeals may function as a guide toward what is metaphysically possible. What I deny is that a mere appeal to the idea that an agent may conceivably perform an act is relevant to the question of what lies in her power under the assumptions (i) that she is nomologically necessitated not to perform that act and (ii) that we are picturing her and her abilities exclusively as part of the natural order.

What does the argument of this section establish? At least I take myself to have shown that Kant's radical interpretation of the idea that we are imperfectly rational agents is not a mere idiosyncrasy. This, I think, follows from the fact that the deflationary interpretations of this idea proposed by Weak Imperfection are unsatisfying. One might, however, worry that the negative argument against Weak Imperfection provided in this section is by itself insufficient to motivate the strong idea that it is a *pervasive* feature of our agency that we find ourselves with genuine options to violate normative laws. I agree: before we can confidently endorse Kant's view, we must identify, in a constructive vein, what *constitutes* the radical imperfection that is inherent in human rationality and agency. I will attempt to do this in section 3 below.

2.3: Imperfect rationality and the deliberative stance

I now want to address a third objection to Kant's argument. The argument relies on the idea that the imperfection in our practical rationality is a fundamental *datum* that any satisfactory account of an objective normative dimension of human agency must be able to accommodate. Now, one might argue that once we expound the reason why the pervasive freedom to violate practical laws seems essential to our self-conception, the following flaw in my argument emerges: the only intuitive basis for thinking of ourselves as imperfectly rational is that we must invoke our freedom to *have done* otherwise to make sense of the idea that we are *responsible* agents. But then, it might be said, the worry about determinism reduces to the implications that determinism has for responsibility, and this shows that I have failed to identify a sense in which determinism threatens our *agency*. A worry about the legitimacy of holding people responsible in a deterministic world arises, one might say, only in a context where we look back at past actions; if we are concerned with the implications that determinism has for agency, the question is how determinism raises a predicament for the forward-looking standpoint of deliberation.²⁷⁹

²⁷⁹ This is a way of understanding Korsgaard's complaint about incompatibilists; 1996b, p. 96: "They aren't talking about a practical problem...but about a theoretical one – that knowledge would show us that we weren't free after all. But how is it supposed to do that? By showing that we could not have done otherwise? That might show that we aren't responsible. But it is a different question whether determinism is a threat to responsibility. Freedom is the capacity to do otherwise, not the capacity to have done otherwise." By contrast, for Kant freedom is *also* the capacity to 'have done otherwise': "All men attribute to themselves freedom of will. Hence come all judgments upon actions that are such that they ought to have been done although they have not been done" (GMS, 4: 455). This is another reason why Kant cannot be read as endorsing the idea (favored, as we saw, by Korsgaard) that our freedom requires only the epistemic indeterminacy of actions: Kant holds that we must

Now, remember my previous point that the experience of temptation is exposed as illusory if it must be conceded that the factors that appear to tempt us cannot really sway us (i.e., if we are necessitated to choose correctly). Notice that this worry has a parallel in cases where we are necessitated to choose incorrectly: in those cases the awareness of our ability to resist temptation is likewise illusory. Combining these diagnoses, we can see that determinism leads to a kind of perversion of the idea that we are, when choosing what to do, subject *both* to the rational influence of norms *and* to conditions that seek to deflect this influence. In a deterministic world, this notion degenerates into the idea that in each concrete deliberative context one of those factors has complete control over the other. This renders concepts such as struggle, temptation or weakness of will generally pointless, for they require that when deciding what to do we have the double option to act on our conception of what we ought to do or to act contrary to this conception. The conviction that struggle, temptation, and weakness of will are essential to our practical experience lies behind the Kantian emphasis on the importance of a real possibility to go either way with regard to objective 'oughts.' Since the experience of temptation, struggle, or weakness of will is characteristic of our deliberative plight, i.e., of the forward-looking stance of agency, we need not appeal to the backward-looking perspective to explain why the awareness of both our rationality and its inherent imperfection are characteristic of our human self-image.

But I do not want to deny that human responsibility and imperfection are closely intertwined. The manner in which a being's will relates to the good and the manner in which a being is responsible constrain one another: in praising God for His actions, one praises His necessarily unfailing good will; but in praising human beings for doing the right thing one praises (at least sometimes) their resistance to the powerful influence of temptation. Now, given this interdependence between responsibility and rational (im)perfection, we should expect that our sense of imputability, just like our sense of rational imperfection, plays a role in forward-looking deliberation. And indeed: it is essential to our self-conception as deliberating agents that we consider ourselves accountable for some of the *future* consequences of choices that we *might* make, namely, praiseworthy for the *intended* consequences of potential good choices and blameworthy for *even the unintended* consequences of potential bad choices. (I acknowledge that if others get into trouble because of my selfish lie, I will not be able to excuse myself by saying that I could not foresee and did not intend this consequence, even if these claims are true.) This sense of future imputability is at least implicit in deliberation. So the features of our sui generis human responsibility for what we end up choosing can be added to the explanation of why our awareness of our freedom to act rightly *or* wrongly is so distinctive of our experience of agency.

(3) Reason and desire: the moral psychology of imperfectly rational agency

As we saw in section 2.2, Kant's incompatibilist argument implies that human agents are always subject to factors whose influence seeks to divert them from the rational influence of practical laws. This suggests a view that I shall refer to as Strong Imperfection: For every (finite) agent, it is always really (metaphysically) possible to violate objective practical norms. Why should we think that this radical conception of human imperfection is correct?

Suppose I see someone lose one thousand dollars. If Strong Imperfection implied that there is an objective practical norm that prescribes returning the money and that everyone would

consider both our future and our past acts free, whereas the idea of epistemic indeterminacy is applicable *only* to future actions. I discuss Kant's conception of moral responsibility in chapter VII.

always be drawn toward keep the money, it would rest on a groundless generalization about human motives. But Kant's idea of Strong Imperfection is different: he holds that there is an objective practical norm that prescribes returning the money *for necessary reasons* and that for every finite agent it is always possible to choose in ways that do not accord with those reasons, namely, on the basis of some contingent desire. This may be (i) a greedy desire to keep the money or (ii) a desire to return the money that is fueled by a fear of being watched, or (iii) a desire to return the money that is fueled by the want that others be happy; etc. If one acts on the basis of (ii) or (iii), one's act of returning the money contains *the letter but not the spirit* of the objective norm (KpV, 5:71-2; RGV, 6:30).

The claim that acting on desire (iii) displays a rationally deficient motivation provokes a common complaint: Kant supposes that acting on the basis of moral reasons is the only way to counteract our tendency to strive for our own pleasure²⁸⁰, and so his view that someone who cares deeply about others is tempted to choose contrary to reason rests on the unacceptably hedonistic view that such a person aims at her own pleasure. Now, even if Kant is committed to such hedonism²⁸¹, this is inessential for the purposes of our discussion. Let us grant that my desire (iii) does not have as its object the pleasure *I get* from making others happy, but *their* happiness. This desire, qua empirical state, depends on my contingent psychological make-up, for instance, on my "love for people and affectionate benevolence" (KpV, 5: 82), emotional states that cause me to identify with the feelings of others and that make it very hard for me to bear their (impending or actual) distress. Objective norms provide reasons for acting that apply to every agent; so the validity and authority of the norm that prescribes returning the money is independent of whether the addressee happens to have benevolent desires. Thus, if my choice of returning the money is based on my affectionate want that others be happy, it is *not* responsive to the objective reasons for returning the money that are represented by practical laws.²⁸² It is important to notice that in this case, my choice reflects a general volitional structure that produces right behavior *only by accident*. This can be confirmed in two ways. First, since my willing is such that I make my affectionate character a condition of my benevolent choices, all it would take for me to choose not to act benevolently is a lapse of affection. This might be due to a change of my hormone level or other factors that dampen my desire to make others happy and thus interfere with the general basis of my choice of benevolent acts. Secondly, even keeping my psychology fixed, my willingness to choose on the basis of my affection will lead to wrong actions in circumstances where making others happy is wrong. Suppose I am a teacher, and a student who has been lazy throughout the semester explains the terrible consequences that failing the class would have for her and asks, with tears in her eyes, for a re-write of the final. If my volitional structure disposes me to act on the basis of my affectionate feelings and my desire that others do not suffer, I will make the wrong choice of granting the re-write.²⁸³

²⁸⁰ See Foot 1972, p. 313; Irwine 1984; Williams 1986a, p. 15, p. 64; and Wood 1984.

²⁸¹ This is denied by Allison 1990, pp. 102-5 and Reath 1989. I think that it is only possible to moderate, but not to deny, the idea that Kant has a hedonistic view of non-moral motives. I discuss this issue in chapter IV, section 2.

²⁸² In Kant's own words: my grounds of choice are, in this case, "only subjectively valid and merely empirical, and [do] not possess the necessity which is conceived in every law, namely, an objective necessity arising from a priori reasons" (KpV, 5: 26).

²⁸³ See (for instance) RGV, 6:30-31 for Kant's emphasis on the fact that if the ground of one's choices is a contingent desire, one's 'outer' (observable) behavior is only *accidentally* in accordance with practical laws. Notice, though, that Kant does *not* hold that if (in the example where I see someone lose one thousand dollars) I return the money to its owner for the right reasons, my moral worth is increased if I must struggle against a desire to keep

Hence, granting normative authority to contingent desires, including benevolent ones, is not a reliable guide to morally required or (even) permissible activity. The other side of this coin is that a free agent who happens to be so psychologically constituted that the suffering of others leaves her emotionally cold can fully comply with the demands of morality (cf. GMS, 4: 398).

Kant's argument for Strong Imperfection can be summed up in the following general terms: (I) Every finite agent is always affected by contingent desires that impel her to 'choose on their behalf', i.e., to make their satisfaction a condition of choice.²⁸⁴ (II) Practical laws give necessary, desire-independent reasons for acting. (III) Thus, it is always possible for every finite agent to choose in ways that defy the spirit of practical laws.²⁸⁵

This argument does not require hedonism about non-moral motives as a hidden premise. It says that *whatever* contingent desires we may have incline us to choose on their behalf and hence impel us not to choose according to reasons whose validity is independent of having those desires.²⁸⁶ That our empirical desires impel us toward their satisfaction seems plausible: to have a desire just is to be drawn, to some extent, to realize its object.

In the remainder of this section, I want to clarify the implications of Strong Imperfection by discussing a number of objections that one might raise against this view. One source of

the money, or decreased if I *also* want to see the owner happy (as long as this want is not my motivating reason for returning the money; see Herman 1996, p. 12, p. 21). But there is (pace Allison 1990, p. 114) something correct about what Henson (1979, p. 48) calls the 'battle citation' model of moral worth: Kant thinks that the specific moral worth of human agency consists in the successful struggle against the temptation *to choose and act for the wrong reasons*. This shows (pace Henson) that a choice cannot be 'overdetermined': I either choose an act for reasons that do or for reasons that do not depend on my desires. There is no middle path here.

²⁸⁴ KpV, 5: 74: "However, we find that our nature as sensible beings is such that the matter of desire (objects of inclination, whether of hope or fear) first presents itself to us; and our pathologically affected self, although it is in its maxims quite unfit for universal legislation, yet, just as if it constituted our entire self, strives to put its pretensions forward first, and to have them acknowledged as the first and original."

²⁸⁵ KpV, 5: 83-4: "...if a rational creature could ever reach this point, that he thoroughly likes to do all moral laws, this would mean that there does not exist in him even the possibility of a desire that would tempt him to deviate from them; for to overcome such a desire always costs the subject some sacrifice and therefore requires self-compulsion, i.e., inward constraint to something that he does not quite like to do; and no creature can ever reach this stage of moral disposition. For, being a creature, and therefore always dependent with respect to what he requires for complete satisfaction, he can never be quite free from desires...and as these rest on physical causes, they can never of themselves coincide with the moral law, the sources of which are quite different; and therefore they make it necessary to found the disposition of one's maxims on moral obligation, not on ready inclination..."

²⁸⁶ Wolf 1980, pp. 152-4 claims that it is *incoherent* to assume that agents who are doing the right thing are not psychologically determined and that they could do otherwise. Her crucial assumption is that an agent can act on the basis of the right reasons *only if* he is psychologically determined (i.e., determined by "his interests" which "are determined by his heredity or environment"). I find this assumption puzzling. It seems based on a subjectivist view of practical reasons which holds that what counts as 'the right reason' for R depends on R's empirically given interests. If one rejects such subjectivism, one can say that an agent who acts on the basis of right reasons is moved by her recognition of reasons whose validity is independent of her contingent interests. Wolf claims that the notion that R had the freedom not to do the right thing implies that R "has the ability to act against everything he believes in and everything he cares about," but this is clearly wrong: R sees alternatives to choosing for the right, necessary reasons precisely because there are things that he 'cares about' by virtue of having empirically given, contingent desires. Hence, pace Lavin 2004, p. 447, the freedom envisaged here does not require that R can "make a choice for any principle": R's alternatives to choosing for desire-independent reasons are delimited by R's contingent desire-base. Wolf also claims (1990, p. 54) that R must represent the right act "as just one alternative among others." But on Kant's account, human agents represent the right act as an alternative that is *rationally superior to all others*, and that is *alone* demanding of moral praise and respect.

resistance toward Strong Imperfection stems from the Aristotelian tradition, in which virtue enables a perfect harmony of reason and desire. The Kantian worry is that this attempt to unify desire and reason must come at the expense of sacrificing the objective, necessary validity of reason's prescriptions: contingently given desires cannot impel us to act for reasons whose validity is independent of our possession of those very desires. Aristotle would respond that both desire and reason aim at *happiness*, which (as the virtuous person realizes) consists in the exercise of the ethical virtues. But this objectivism about what we empirically desire (happiness) requires that there is a natural *telos* for human activity which guarantees that *only* someone who exercises the virtues can find a satisfaction of the desires imposed on her by her empirical nature.²⁸⁷ Many (though not all) Neo-Aristotelians reject this natural teleology; they secure the virtuous person's immunity to temptation by making what practical reason deems worthy of pursuit dependent on an agent's contingent desire-base.²⁸⁸ Thus, rejecting Strong Imperfection comes at great cost: it requires *either* a value-laden, pre-modern view of nature *or* a sacrifice of the idea that there are laws of practical reason. This supports the Kantian point that contingent desires cannot constitute grounds of choices that respect the spirit of necessary laws of reason.

Now, even if one grants that Kant's conception of virtue qua continence or "moral disposition [Gesinnung] in struggle" (KpV, 5: 84) is, for the reasons mentioned, preferable to the Aristotelian conception of virtue, one might still worry that Strong Imperfection is implausible on phenomenological grounds: it appears to imply the ubiquitous presence of temptation and struggle in ordinary everyday action, and this seems to involve a gross distortion of the phenomena. But I do not think that Strong Imperfection is guilty of such distortion: the idea that it is always possible for every (finite) agent to choose in ways that defy the spirit of practical laws does *not* imply that each and every one of our choices is accompanied by the conscious awareness of temptation. This is so for a simple reason: becoming aware of our practical imperfection in the form of temptation requires a reflective awareness of the two competing factors in our motivational economy that may serve as grounds of choice, namely, contingent

²⁸⁷ I am alluding here to Aristotle's 'function argument,' which aims to show that happiness (the human good) consists in the exercise of the ethical virtues as a matter of natural fact. Aristotle and Kant agree that empirical desires aim at 'happiness', but Kant rejects Aristotle's teleology and objectivism about happiness: for Kant, happiness relates to the pleasure we get from satisfying our given desires, *whatever they are*. This disagreement between Kant and Aristotle is not merely verbal: Aristotelian 'eudaimonia' is the ideal of a life that is pleasant on the whole ("the truly happy man will also live most pleasantly," EE, 1249a18; compare EN, 1152b5-8), and for Aristotle our conception of the good is grounded in what we happen to find pleasant (see EN, 1104b9 ff.; EE, 1227a37 ff.). Aristotle claims (EN, 1166b13-29), while Kant denies (KpV, 5: 114), that a vicious agent is confused about what gives her lasting pleasure and cannot find her life pleasant on the whole. Likewise, Aristotle claims, while Kant denies, that an agent who (continuously) exercises the virtues is *bound to* enjoy her life.

²⁸⁸ Foot 1972, Williams 1986 and Anscombe 1958 all reject the idea that there can be necessary *laws* of reason. By contrast, McDowell 1978 treats moral requirements as categorical imperatives that, in the case of the virtuous agent, completely 'silence' competing motives. But this is puzzling: the Aristotelian virtuous agent has no competing motives that could be silenced, and hence (as Anscombe notices) such an agent does not represent ethical norms to herself as *laws* that constrain her activity (i.e., she faces no 'oughts'). McDowell may think that the virtuous agent has motives that *would* tempt her if it *weren't* for the influence of ethical norms. But since the virtuous agent is *always* influenced by ethical norms, this counterfactual is empty. So McDowell's picture collapses into the idea that all desires of the virtuous agent aim at the exercise of virtue. This raises my question of how contingently given desires can impel one to act for desire-independent reasons. I cannot consider here how Neo-Aristotelians such as Foot (in her *later* work) or Michael Thompson would answer this question, but I do not think that they achieve the kind of harmony of reason and desire that would falsify Strong Imperfection.

desires and desire-independent reasons. But deliberative attention to, or determination of, our reasons for acting is not a typical component of intentional activity: we make many of our everyday choices unreflectively, by taking for granted that certain things rather than others are worth doing. Hence, from the fact that it is always possible for finite agents to choose contrary to the spirit of practical laws it does not follow that this possibility is always or even typically an object of conscious awareness, in the form of temptation or struggle.

For Kant, one context in which we *inevitably* become acutely aware of this possibility is that in which we choose what he calls maxims. Maxims are general principles of willing that exhibit stable volitional structures, that is, policies or dispositions to accept (without further reflection) certain considerations as reason-giving and as grounds of particular choices. Kant's view that maxims play a crucial role in our intentional activity derives its plausibility from the combination of the aforementioned fact that most of our everyday choices are not preceded or accompanied by explicit deliberation and the further fact that we assume that unreflective choices at least *tend to* reveal what the agent sees as her reasons for doing one thing rather than another. This warrants the attribution of general principles of choice that underlie our choices of specific courses of action. The reflective adoption of these general principles affords us with a vivid awareness of the fact that we may go 'either way' with regard to objective norms, because it forces us to confront the question of what place we assign to moral principles and to the pursuit of happiness in our life – that is, whether we subordinate the bidding of our contingent desires to the weight of necessary reasons or vice versa in the formation of our general policies. Notice, however, that the idea that our awareness of our double option to comply with or violate practical is especially prominent in contexts where we choose general policies does *not* preclude that an awareness of this double option (and hence an awareness of being tempted to defy the spirit of normative laws) may also arise with regard to individual acts: for instance, and characteristically, acts that would signal a weak-willed departure from one's lawful principles of willing.

Hence, Strong Imperfection is compatible with the intuition that struggle and temptation do not (for better or worse) abound in our everyday activity. The possibility that we may defy the spirit of laws of reasons may not typically occur to us – because certain unreflective habits of choice have become second nature to us, or (to mention another prominent manifestation of our practical imperfection) because we succumb to our proneness to self-deception about our motives. In this latter case, we take ourselves to be acting on the basis of universally valid, necessary reasons whereas in fact some contingent desire constitutes the ground of her choices. One can suspect that a volitional structure of this kind is the source of a great many injustices, for while agency based on contingent desires produces (as we saw) morally permissible behavior only by accident, an agent who is self-deceived about her motives will proceed with great confidence in the universalizability of her motives and the rightness of her dispositions.

Now, one might suggest that the appeal to self-deception provides a welcome ground for moderating Strong Imperfection, in the following manner: instead of positing the *metaphysical* possibility that agents may always choose contrary to the spirit of practical laws, we should make use of the weaker idea that it is always *epistemically* possible for a finite agent to choose incorrectly, in the sense that she can never be certain that she actually succeeds in determining herself to act on the basis of necessary, desire-independent reasons. Kant clearly accepts this epistemic idea (cf. *GMS*, 4: 419); but I think that he accepts it only as the epistemic upshot or consequence of the robustly metaphysical idea inherent in Strong Imperfection. That is: what

grounds the ubiquitous doubt about whether we succeed in respecting the spirit of practical laws is precisely the ubiquitous metaphysical possibility that we may fail to respect the spirit of practical laws. Unless one had reason to take seriously this metaphysical possibility, one's epistemic uncertainty about the adequacy of one's motives would be unmotivated. One may object that this reasoning is fallacious: even if it is *not* always metaphysically possible for finite agents to act on the basis of inadequate motives, they can always doubt the adequacy of their motives because they do not know whether or not the metaphysical possibility that one may choose contrary to the spirit of objective norms obtains in a given case. But this strikes me as a rather artificial position: if one concedes that we are never in a position to rule out whether this possibility obtains, why would one insist that it is implausibly strong to suppose that this possibility always obtains? It would seem that whatever grounds one has for finding Strong Imperfection too extreme or implausibly strong should also count against the claim that it always makes perfect epistemic sense for us to suppose that our motives might fail to display respect for practical laws. And indeed, our epistemic grounds for supposing that our motives might be inadequate can only relate to the metaphysical fact that for any choice that may be based on necessary, desire-independent reasons, there is a (at least one) empirical desire that may also (by way of temptation or self-deception) function as the ground of choice.²⁸⁹ That there is such a desire is guaranteed simply by the fact that we are dependent beings, and as such are always under the influence of (contingent, empirical) motives that impel us to act in a spirit that cannot coincide with the spirit of (necessary, a priori) laws of practical reason.

I want to consider a final worry about Strong Imperfection. This worry derives from a *rationalistic view of desire* according to which desiring as such involves seeing a reason.²⁹⁰ A proponent of such a view might argue that since desires are themselves infused with rationality, they cannot be pitted against reason, and (likewise) that choices based on desire cannot be pitted against choices based on reason, in the manner envisaged by Strong Imperfection. The most we can say on the rationalistic view is that there might be a difference in degree between rational motivation based on given desires and rational motivation based on normative principles (desires representing weaker or less weighty reasons for acting); but on a rationalistic view there is nothing that prevents us from supposing that our given desires might represent the very same reasons that also inform our awareness of normative principles.

One problem with the rationalistic account of desire is that it conflicts with phenomena such as desiring acts one deems contrary to reason. Suppose I have done something bad, and now I feel tempted to tell a lie so that I can pin the blame on somebody else. The proponent of the rationalistic account must interpret this case as implying that qua desiring, I see at least a weak *pro tanto* reason to lie. But in accepting that my lie is prohibited by an unconditional law of practical reason, I reject the idea that the option of lying is supported by any reason whatsoever: I accept that the bidding of my desire to escape reproaches is rendered, rationally and normatively speaking, *null and void* by the demand not to treat others as a means to my personal gratification. The 'market price' of my personal gratification cannot be rationally weighed against

²⁸⁹ This point bears an interesting parallel to the argument of section 1: Kant does accept that our (noumenal) acts are epistemically open to us, but this is entirely derivative of his acceptance of the fact that our noumenal acts are metaphysically open (and it is completely independent of the further fact that he also accepts that our phenomenal acts are epistemically open to us given our lack of insight into the psychological space of causes).

²⁹⁰ For such rationalism, see Scanlon 2000, p. 20, p. 38, p. 40; and Stroud 2011, chapter 4.

the value of human dignity.²⁹¹ The proponent of the rationalistic account may respond that this implies that I must think of my desire as an unintelligible 'urge' devoid of intentional content.²⁹² But this does not seem correct. My desire to lie represents the prospect of avoiding the distinctive unpleasantness associated with the experience of being an object of indignation; this representation makes telling the lie a tempting option, independently of an implausibly strong commitment to the idea that there is always a reason to avoid the unpleasant.²⁹³

The rationalistic account of desiring faces a further problem. Proponents of this account tend to reject the claim that having a desire as such implies having a reason.²⁹⁴ But this commitment sits uneasily with the idea that desires as such represent reasons: if there is no intrinsic link between *having* a desire and *having* a reason, this fact should be accessible to deliberation; and this seems to preclude an intrinsic link between *having* a desire and *seeing* a reason. The rationalist might respond that having a desire, as such, involves grasping the *appearance* of a reason, whereas reflection is needed to determine whether we in fact have reason to act on our desires. This is a consistent position, but it does have two rather unappealing implications. First, on this view, the very content and phenomenology of desiring involves a kind of delusion: states that intrinsically lack normative authority intrinsically appear to us as staking a normative claim. Second, the rationalistic view threatens to collapse the difference between the passive possession of a desire that gives us no reason to act and the active endorsement of such a desire as reason-giving that occurs in a false practical judgment. Not all bad choices are weak-willed: a bad choice may be based on the false representation that something is worth doing. Here it seems that the choice is based on bad or incorrect reflection which is responsible for the non-veridical appearance of normativity; but on the rationalistic picture there is no need to posit an act of incorrect reflection and judgment, for the (false) appearance of normativity that underlies the choice is already inherent in the very possession of the relevant desire. We may still, on the rationalist picture, fault the agent for failing to check on the veridicality of the appearances, but this criticism is considerably thinner, and less serious, than the criticism that she reached an erroneous judgment (as the bearer of normative content); correspondingly, on the rationalist story the responsibility of passively given desires for bad action is increased considerably, and blame is at least some extent diverted from the agent's free activity.

For Kant, having a desire is an entirely passive matter that does not per se (apart from an active endorsement of the desire as worth satisfying) involve the representation of a normative reason: such representation is always within the scope of rational agency and responsibility.²⁹⁵ Hence, for Kant (who also rejects the subjectivist link between desiring and *having* a reason), our practical experience is shaped by the gap between our awareness of desires and our awareness of

²⁹¹ See GMS, 4: 434-5, and my discussion in chapter IV, sections 1 and 2, concerning the two distinct values represented (respectively) by categorical and hypothetical judgments.

²⁹² See Quinn 1994; Scanlon 2000, p. 38 ff.; and Stroud 2011, p. 109 ff.

²⁹³ Schapiro 2009 provides a forceful critique of rationalism about desiring. However, I am skeptical about her view that desires represent an action as *to-be-done* in virtue of features that make doing it look good (ibid., pp. 252-3). The imperative form is indicative of a special kind of authority that we should, with Kant, reserve for prescriptions of reason. Schapiro rejects the rationalistic idea that to desire *x* is to see oneself justified in the pursuit of *x*; but the content she assigns to desires (an act is *to-be-done* in virtue of its attractiveness) seems cognitively thick and authoritative enough to lend itself to justificatory thoughts.

²⁹⁴ See Scanlon's discussion (2000, pp. 363-74) of William's desire-based theory.

²⁹⁵ To endorse a desire as reason-giving is to adopt a general rule that deems the desire worth satisfying, and which underlies particular (unreflective) choices. See RGV, 6: 24; and Allison's 1990, pp. 39-40 'incorporation thesis'.

normativity. This, I suggest, is the plausible basis of his appeal to Strong Imperfection.

4: Incompatibilism and the metaphysics of law-governed agency

I conclude that Kant's incompatibilism, far from being "a mere dogmatic attachment"²⁹⁶, has a deep foundation in our self-conception. We, as finite rational agents, are constantly under the influence of contingent desires that impel us to choose contrary to the influence of necessary practical laws. This provides for the distinctive experience of human agency, which is inextricably (at least in the context of deliberative reflection about what to do) tied to the awareness of a pervasive, not merely epistemic possibility to go either way with regard to objective norms. Both the recognition of the validity of normative laws and the recognition of the attractiveness of options other than complying with those laws point us toward possibilities that do not merely exist for all we can know. That is, they point us toward the deep metaphysical idea that our ability to act on the basis of objective reasons has a *shadow*, our propensity to choose contrary to the weight of such reasons, that perpetually hovers over our acts of practical self-governance. This accounts for a kind of freedom of will that is shared neither by God nor by donkeys nor by computers.²⁹⁷ Such freedom is presupposed by the formula in which objective normative laws are represented to imperfectly rational agents: 'ought.' Determinism, understood as a universal metaphysical truth, would deprive us of such freedom. Hence, it would deprive objective 'oughts' of their point; and this would strip our actions of the normative dimension that they owe to being governed by meaningful objective 'oughts.'

This line of thought strictly presupposes the existence of norms whose reason-giving force is categorically at odds with the influence of desire and that have a pervasive impact on our practical experience.²⁹⁸ For Kant, only the norms of morality *necessarily* satisfy these conditions. What an agent has *non-moral* reason to do depends on her contingent desires, namely, on what happens to give her lasting contentment (i.e., happiness).²⁹⁹ An appeal to subjective norms of prudence cannot conclusively support Kant's incompatibilism, because awareness of such norms is not *essentially* coupled with an awareness of temptation. True, in the actual world, temptations to act contrary to norms of prudence abound, and we are in need of the help of practical reason against the influence of desires whose satisfaction would make us unhappy. But it is conceivable

²⁹⁶ Ameriks 2000, p. 227.

²⁹⁷ I am not proposing that on Kant's account our freedom of will can be *theoretically defined* by reference to our propensity to act wrongly: for Kant, a non-empirical capacity (such as freedom) cannot be theoretically understood. This is clear from an important passage (MS, 6: 226-7) in the *Metaphysics of Morals*. But this passage does not negate the idea that Kant accepts that the tendency to choose incorrectly is characteristic (albeit not in a way that supports a strict definition) of our freedom. What Kant rejects in this passage is the idea that we have a firm theoretical grip on our freedom to choose incorrectly on the basis of our *observation* of wrong actions. Kant himself has a priori grounds (see RGV, 6:35-6) for accepting that the free choices of finite rational beings take place against the background of affection by desires that exert a tempting influence on the will and hence (partly) underlie our propensity to choose for the wrong reasons. It is *only* by taking seriously the omnipresence of this 'shadow' of our power to choose correctly that one can appreciate Kant's idea that moral motivation involves both humiliation by and respect for the moral law (KpV, 5: 74-5).

²⁹⁸ For an argument against merely desire-based conceptions of practical normativity, see chapter V, section 2.

²⁹⁹ See KpV, 5: 25-6. We must not confuse Kant's conception of non-moral normative reasons with contemporary desire-based accounts. For Kant, it is an objective truth of practical reason that every (finite) agent has non-moral reason to do what would make her happy, pending the moral permissibility of her act; the subjectivity of such reasons derives from the fact that what makes people happy depends on their contingent psychology. See chapters IV and V for extensive discussion of these points about hypothetical imperatives.

that our recognition of the means to our happiness might infallibly give rise to an unailing motive to take those means. In this case, our non-moral choices would need no assistance from practical reason, the discovery of means-end relations being a task for theoretical reason³⁰⁰; we would never face a temptation to make imprudent choices; and thus rules of prudence would not strike us in the shape of 'oughts.' By contrast, no contingent empirical desire can conceivably replace the impact that practical reason has on a choice that is made for necessary, desire-independent reasons; a finite rational agent cannot conceivably fail to find herself with options to choose contrary to the weight of objective reasons; and thus the objective norms of morality necessarily strike imperfectly rational beings in the shape of 'oughts.'

Thus, if we focus our attention on *non-moral agency*, the compatibilist position that Kant dismisses with regard to moral agency is not necessarily unsatisfactory. Compatibilism seems unsatisfying with regard to our non-moral agency in the actual world, because our actual experience of non-moral agency is fraught with temptation and weakness of will. But we can conceive of a world in which finite agents are infallibly motivated to do what makes them happy and never feel tempted to act imprudently. If we make the further supposition that in this world morality is an illusion, so that no unconditional requirements of reason (objective practical laws) exist, we have imagined a world in which compatibilism yields the correct analysis of agency.³⁰¹

I want to point out an important consequence this line of thought has for the metaphysics of normativity. In this essay, I have provided a foundation for tying the law-governed agency of finitely rational beings to a pervasive possibility of a violation of normative laws. This foundation has two cornerstones: first, such beings are susceptible to a ubiquitous propensity to choose on behalf of their contingent desires; second, they are subject to 'oughts' that represent desire-independent reasons for acting. This foundation collapses in the case of divine agency, because the first cornerstone is removed here. God's guidance by reason is not tied to the possibility of resisting reason because there is no metaphysical basis for such resistance: God has

³⁰⁰ For Kant's claim that the rules of prudence are established by theoretical reasoning, see KU, 5: 172-3; EEKU, 20: 196-7. One might object that hypothetical imperatives presuppose categorical imperatives, since one needs non-instrumental reasons to take means toward one's ends (see Hampton 1998; Korsgaard 1997). Kant himself never argues from the hypothetical to the categorical imperative, and he states that the existence of hypothetical norms does not suffice to verify the existence of categorical norms (GMS, 4: 419). In response to Hampton's and Korsgaard's objection, we must notice that while all hypothetical imperatives give instrumental reasons insofar as they specify means toward happiness, happiness itself is *not* a means to something further. So the appeal to happiness provides a non-instrumental ground of instrumental reasons (and an agent who always reliably desires what in fact makes her happy needs no assistance from practical reason). We must not confuse here the fact that the end of happiness has a *non-instrumental* status with the idea that it is an *unconditional* good: while we desire happiness for its own sake, the value of happiness (as the one and only non-moral value or 'Wohl' we are concerned with as dependent, finite beings who are afflicted with a sensible nature) is subordinate to the value of morality (GMS, 4: 393). Thus, happiness, qua merely conditional good, cannot ground unconditional normative demands. See chapter IV, section 2, and chapter V, section 3 for further discussion.

³⁰¹ Kant himself imagines such a world in the Groundwork (GMS, 4: 395), when he invokes a scenario in which the *only* purpose of finite beings is the achievement of their own happiness. He says that in such a case it would be preferable if practical reason did not interfere with our natural instincts; and he emphasizes (KpV, 5: 61-2) that as things stand in the actual world (where we lack instincts that are infallibly attuned to our sensible 'welfare'), practical reason in its non-moral function must serve the role that instincts serve for animals. It is only with regard to morality that practical reason plays a role that is *necessarily irreducible* to the influence of empirically given desire; consequently it is only the consciousness of objective norms of morality that *intrinsically* involves the awareness of free will (KpV, 5: 29-30), and freedom is a ground of the possibility of categorical imperatives only (MS, 6: 222).

no finite, empirical nature that afflicts Him with desires that account for the presence of options other than compliance with laws of reason. This foundation *also* collapses in the above scenario where there are no objective norms and our empirical nature infallibly directs us to attain the ends that it imposes. In this case, the second cornerstone is removed: here we lack a faculty of practical reason that is an autonomous source of desire-independent ends and laws in relation to which a choice based on a contingent desire could be deemed erroneous or improper.

My argument thus suggests a response to Christine Korsgaard's claim that even a perfectly rational being "cannot be conceived as guided by reason unless it is conceived as capable of resisting reason."³⁰² My response starts from the idea that one cannot dogmatically presuppose, without justification, that norm-governed agency *is* essentially tied to the possibility of practical error. The justification I have provided for the intuition that norm-governed agency requires the possibility of resisting reason directs us to features that are specific to the law-governed agency exhibited by *imperfectly* rational beings. Hence, this intuition stands, *pace* Korsgaard (unless she can offer an alternative justification³⁰³), entirely unsupported in the case of the law-governed agency exhibited by a *perfectly* rational being.³⁰⁴

³⁰² Korsgaard 1997, p. 240. Compare her 1996b, p. 161: "there is no normativity if you cannot be wrong."

³⁰³ Korsgaard suggests (1997, p. 240) that we can only *describe* the acts of a divine agent. The point here seems to be that with respect to a being that cannot fail to comply with normative rules, we lose our grip on the very idea of normative guidance and must consider such a being as a kind of *automaton* (see Lavin 2004, pp. 443-4 for this diagnosis, which he does not endorse). But we have many resources for distinguishing God from automata. A mechanism is entirely determined by *non-rational* elements, governed by physical laws, and programmed to behave as it does by an external designer. God, by contrast, is not 'set up' in any way by external causes and entirely removed from the nexus of non-rational (material) determining factors. Hence, nothing stands in the way of the diagnosis that God (unlike a mechanism) is infallibly guided by His recognition of the weight of reasons.

³⁰⁴ This conclusion is also reached by Lavin's very instructive 2004 discussion. But there are crucial differences between our respective arguments. Lavin operates (cf. footnote 277 above) with a deflationary, de-radicalized conception of the possibility of error inherent in the idea of imperfectly rational agency. His argument against Korsgaard consists (a) in the claim that we are not forced to extend this conception to God and (b) in the allegation that so extending it is incompatible with 'constitutivism.' By contrast, my argument provides a unified explanation for why norm-governed imperfectly rational agency involves a radical, metaphysical possibility of error *and* for why norm-governed divine agency must be exempt from this 'error constraint.'

Chapter VII: Possibility and Responsibility

Kant never put forward a systematic account of moral responsibility, that is, of what our judgments of responsibility require for their legitimacy. His relevant remarks are often mere corollaries of his discussion of freedom. In this paper, I aim to reconstruct Kant's conception of freedom as a systematic libertarian position. I first present and defend what I take to be Kant's argument for incompatibilism about responsibility. I then raise and try to resolve a worry about the coherence of this argument with his positive, libertarian view, before sketching what this view implies in concreto for our practice of holding people responsible. This concludes my discussion of the foundations of Kant's incompatibilism about practical freedom.

1: Kant's incompatibilism about moral responsibility

Kant's conception of moral responsibility must be considered in relation to his conception of freedom, and hence in relation to the general framework of transcendental idealism that, Kant believes, uniquely allows for the belief in freedom. We can only consider people morally responsible if we can regard their will as transcendently free; hence Kant's conception of moral responsibility inherits all the features of Kant's doctrine of freedom. This means, in particular, that Kant's position relies on the idea (for which his idealism is supposed to make room) that we can attribute our actions to an unconditioned causality that operates outside the temporal, deterministic order of nature – a notion that is coherent but theoretically fruitless *by Kant's very own standards*. Hence, the attractiveness of Kant's conception of moral responsibility depends to a large extent on the untenability of alternative models, in particular of models that rely on a metaphysically modest, compatibilist conception of responsibility. It is only if such compatibilist proposals fail that Kant's libertarian account can seem attractive despite the fact that it draws and depends on ideas that are theoretically (e.g., explanatorily) fruitless. So, Kant's account of responsibility must rely on an argument against compatibilism. (This is an application of a general point, which I expound in chapter II, to the issue of responsibility.)

I attribute the following incompatibilist argument to Kant: (1) To judge that someone is responsible for her action is to imply that she could have performed a different action than the one she in fact performed. (2) If there were only the deterministic causality of nature (*in short, if (d) were true*), no agent could have acted differently than she in fact did. (3) If (d) were true, judgments of responsibility would be illegitimate because they would imply something false.³⁰⁵

³⁰⁵ For premise (1), see CPR, A 555-6/B 583-4: "Our blame is based on a law of reason whereby we regard reason as a cause that irrespective of all the above mentioned empirical conditions could have determined, and ought to have determined, the agent to act otherwise. (...) ...we say that in spite of his whole previous course of life the agent could have refrained from lying..."; and KpV, 5:95, where Kant says that a moral assessment of a theft requires "that [the theft] could have been omitted, because the law says that it ought to have been omitted". For premise (2), i.e., for the idea that if our actions were solely subject to natural causality, the freedom required for responsibility would be impossible, see CPR, A 534-38/B 562-66 and KpV, 5: 96-7.

I disagree with Korsgaard's claim (1996, p. 197 ff.) that Kant adopts a 'practical' as opposed to a 'theoretical' conception of responsibility (for a similar 'non-factual' interpretation see Reath 2006, p. 251). Korsgaard says that a 'theoretical' conception holds "that it is a fact...that [a person] could have acted otherwise"; the above quotes show that Kant's conception of responsibility *is* 'theoretical' in this sense. Korsgaard might respond that for Kant the claim that a person could have acted otherwise carries no factual implications because it is made from a practical standpoint (*ibid.*, pp. 200-5). But this response is unsatisfying. It gives rise to the general problem (which I discuss at length in chapter II) that if Kant's conception of responsibility were wholly devoid of 'factual' assumptions about our independence from natural necessitation, it would be available to a

In this first section, I want to discuss the cogency of this rather straightforward argument in relation to some common compatibilist objections. My aim is not to give a conclusive defense of Kant's incompatibilist position, but rather to show that it is not (as is often claimed) based on mere dogmatism, and hence that it should be taken seriously on philosophical grounds.

As far as (1) is concerned, we can distinguish between three different types of compatibilism that reject this premise. First, there is a *consequentialist* compatibilism which says that we can praise and blame agents for past actions if doing so increases the likelihood that they will behave as we want them to behave in the future (blame, in particular, serves the function of deterrence).³⁰⁶ Kant would surely protest that this is a rather unappealing *instrumentalism*: it implies that human agents are responsible in much the same way as dogs and donkeys are, and that when we hold someone responsible we view her as a target of causal manipulation, a means toward our desire-satisfaction. Moreover, consequentialist compatibilism ignores that our actual practice of holding people responsible is backward-looking rather than forward-looking: when we hold someone responsible we focus on what she *has done* compared to what she ought to have done, quite irrespective of what we predict or expect concerning her future behavior. (Even if we knew that our blame or praise has no effect on a rational agent's future behavior, we would still think that our blame or praise is perfectly legitimate.) Hence, the consequentialist rejection of (1) seems to invoke a kind of revisionism about our actual practices of holding people morally responsible.³⁰⁷ But Kant seeks to defend our actual moral practices.

A second objection to (1) is suggested by Harry Frankfurt's counterexamples.³⁰⁸ Suppose A kills B as a result of her own volition. If so, Frankfurt argues, the fact that A would have been mentally coerced by C to decide to kill B if C had foreseen that A will *not* voluntarily decide to kill B is irrelevant; A's lack of possibility not to decide to kill B does not show that A is not responsible for killing B. To see how Kant would respond to this argument, notice first that for Kant choices are 'inner' actions.³⁰⁹ Now, it seems that Frankfurt's argument presupposes that A's free decision to kill B is *the same inner act* as the decision to kill B that results from C's mental coercion. But a bad choice one freely makes seems to differ intrinsically, by virtue of reflecting the agent's moral character and her failure to acknowledge the force of moral reasons, from a choice that results from mental coercion or manipulation. Correspondingly, killing someone out of one's own, free volition and killing someone as a result of a volition that results from coercion are *different acts*. Hence, Kant could reply to Frankfurt that *the act* that we hold A responsible for is one that A was in a position to omit. Thus the 'principle of alternative possibilities' that Kant's

naturalistically inclined transcendental realist. Moreover, it is elusive how the claim, 'she could have acted otherwise' could be interpreted in a 'merely practical' manner that cancels its apparent factual implications. A 'merely practical' conception of responsibility, whatever that would be, should look to avoid the principle of alternative possibilities altogether. That this is in fact Korsgaard's own, compatibilist position is clear from what she says elsewhere (1996b, p. 97). But this must be kept separate from Kant's own view. Similarly, there is (pace Reath 2006, p. 253) *no* textual basis for asserting that for Kant freedom and responsibility play a role only in *prospective deliberation*. "All men attribute to themselves freedom of will. Hence come all judgments upon actions that are such that they ought to have been done although they have not been done" (GMS, 4: 455).

³⁰⁶ See Bentham 2007; Dennett 1984; Smart 2003.

³⁰⁷ See Wallace 1994, pp. 55-9 for an elaboration of these and similar kinds of objections.

³⁰⁸ See Frankfurt 2003b.

³⁰⁹ See MM, 6: 214 for Kant's distinction between 'inner' and 'outer' actions. See also my chapter III, section 3; and compare Korsgaard 1996a, p. 179: "Making something your end is a kind of internal action".

incompatibilism requires seems to survive Frankfurt's attack.³¹⁰

A third argument against (1) has been proposed by Jay Wallace, who tries to find a principled basis for the irrelevance of alternative possibilities to moral responsibility in our moral practices. He rightly presupposes that we hold agents responsible on the basis of a belief that they have violated a moral obligation. He then argues as follows: (I) Incompatibilism about responsibility rests on the idea that our reasons for *excusing* and *exempting* people imply the principle of alternative possibilities and that these reasons would generalize if (d) were true.³¹¹ (II) But ordinary *excuses* (such as not blaming a chained man for not trying to save a drowning child) do not appeal to a lack of alternative possibilities but to "the absence of a culpable quality of will."³¹² (III) Ordinary *exemptions* presuppose only that an agent (say, an insane person) lacks the general power for responding to moral reasons, not that he could have exercised these powers differently.³¹³ (IV) Thus, responsibility does not require alternative possibilities.

There is an important point of agreement here between Wallace and Kant: for anyone, like Kant, who supposes that what we hold people responsible for is their volitional structure ('the quality of their will'), it is hard to see how the question of whether an agent has alternative possibilities at the 'outer', extra-volitional level can be decisive for the question of whether the agent is praise – or blameworthy. (I will elaborate on this point in section 3 below.)

But this does not show that alternative possibilities are not required at the *inner or volitional* level. Consider (II). The notion of a 'quality of will' applies to chosen volitional states, i.e., to intentions or maxims of varying generality that can be said to be required or permitted by moral obligations.³¹⁴ As we saw, for Kant making a choice is an inner action, a result of the exercise of an agent's volitional capacities; and we surely presume that it was in a responsible agent's power *not to make* a morally bad choice. Wallace would insist that we can understand this notion of power without any appeal to alternative possibilities (namely, merely in terms of the 'general capacity' to choose rightly), but this is not something that his argument concerning (II) *establishes*. So, the description of our ordinary practice of excuses is at least consistent with the idea that we presuppose that a responsible agent had alternative possibilities (at the level of choice of qualities of will). Moreover, there is a reason to think that the centrality of the notion of a quality of the will to our practice of responsibility does imply our commitment to alternative possibilities. Notice (this is surely a point that Kant would put great emphasis on) that we do not usually apply this notion to emotions and feelings whose presence or absence is independent of an agent's reflection on norms³¹⁵: for instance, A is not culpable for experiencing a feeling of gratification when he hears about the demise of B who mistreated him; but A is culpable for his chosen volition to bring about B's demise or to omit interfering with the conditions of B's demise (if he learned about them).³¹⁶ The most natural way to interpret our practice of holding people

³¹⁰ For a different way of framing these and related criticisms of Frankfurt, see Widerker 2003.

³¹¹ Wallace 1994, pp. 114-7.

³¹² Wallace 1994, p. 151.

³¹³ Wallace 1994, p. 190.

³¹⁴ See Wallace 1994, pp. 141-2.

³¹⁵ This point is endorsed by Wallace 1994, pp. 133-4.

³¹⁶ This is not to say that facts about A's feelings and emotions have no implications whatsoever concerning her moral character, and there may even be a requirement on A to 'work on her emotions' in such a way that she no longer feels gratified about other's demise. But at the very least – and this is the essential point here – we must admit that our emotions and feelings are not subject to our direct volitional control in the way that our choices are; and we must also admit that there is no guarantee or even a specifiable likelihood of success even for

responsible for their volitions but not for their feelings is, it seems to me, by appealing to our conviction that people lack alternative possibilities with regard to how they feel, but not with regard to how they choose.

Now with regard to (III), I agree that we appeal, in exempting (say) a clinically insane agent A, to A's lack of the general power to respond to moral reasons. But we might do so *precisely because* A's lack of this power implies that A is never in a position to exercise it: our knowledge that A lacks, as a matter of general fact, the power to respond to reasons preempts the question of whether A was able to exercise this power on a particular occasion. So Wallace's description of our practice of exemptions is, again, at least compatible with the notion that this practice is committed to the principle of alternative possibilities; and this principle can even be cited to explain his description. Moreover, there is a reason for thinking that we exempt agents by appeal to the absence of the general power of responding to reasons *because, and only because*, this absence implies the impossibility to exercise that power: while the absence of this power is sufficient to show that the agent is not responsible for her actions, the mere presence of this power seems insufficient for responsible agency. An example I used in the preceding chapter (section 2.2) is helpful here: we ascribe the general capacity to walk to a man who is locked in iron chains and thus cannot walk. This shows, at least, that we often find it important to distinguish between the ascription of general capacities and the possibility to exercise those capacities. Now, we can imagine many scenarios in which we ascribe the general capacity of recognizing and responding to reasons to people, but do not consider them responsible agents because they cannot exercise this capacity: consider Sleeping Beauty, or Aristotle's case of a virtuous agent who sleeps through all of her life but is not 'happy' (precisely because she has *merely* the general 'hexis' for virtuous activity); or, more realistically, a person who is viciously drunk or under the influence of hypnosis. More generally, I find it a bit hard to see how Wallace's claim that "what matters is not the exercise of the general powers of reflective self-control, but the possession of such powers"³¹⁷ could be affirmed by someone who considers herself a rational, responsible agent. Our possession of these powers has significance for us, it seems, only insofar as we can exercise them in concrete acts of choice: these, after all, are the *objects* of judgments of responsibility.³¹⁸ Why should my possession of a general capacity of reflective self-control matter to me for any reason other than that I can exercise it in one way or another?

In sum, we can see that Wallace's argument against the significance of alternative possibilities for responsibility does not succeed, for he does not show that our practices of exemption and excuses do not require alternative possibilities (at the inner or volitional level). Moreover, we noticed some considerations that suggest (if not conclusively) that our practices of

someone who tries very hard to 'manage' her emotional states.

³¹⁷ Wallace 1994, p. 90.

³¹⁸ It is clear that incompatibilists generally worry about the exercise rather than the possession of general abilities. See Wolf 2003, p. 387: "according to our commonsense understandings, having [an] ability is one thing and exercising it is another"; compare D. Locke 1973-4, p. 179, and van Inwagen 1983, p. 13. Wallace suggests that this focus on the exercise of general capacities is a mistake (1994, p.183). But it is difficult to see how the idea of a general capacity could have such absolute priority over the idea of the exercise of this capacity on concrete occasions. Wallace himself holds (p. 131) that "*what* one is obligated to do must be the sort of thing that could be motivated by one's grasp of the reasons expressed in moral principles". Why would one think that the 'could' in this sentence is only that of 'general power'? The *only* sort of thing that is, or is not, motivated by a grasp of moral reasons is a concrete choice that results from a concrete exercise of volitional powers, an exercise that respects or defies the weight of moral reasons.

exemption and excuses involve a positive commitment to alternative possibilities.

Thus, premise (1) of Kant's incompatibilist argument appears to be defensible against the most important compatibilist objections. One might insist, however, that Kant needs a more systematic positive motivation of the idea that the notion of alternative possibilities is integral to our self-conception as responsible agents. In response, it can be noticed that Kant frequently illustrates (1) by appeal to the connection between moral responsibility and moral duty. We hold agents morally responsible for violating or complying with a moral duty; the notion that some action falls under a moral duty is inextricably linked to the notion that performing or not performing the action was in the agent's power (KpV, 5:95); it follows that holding an agent morally responsible for an action is inextricably linked to the notion that it was in her power to perform or not to perform the action. The deeper point underlying this line of thought is one that I expounded (and defended) at length in the preceding chapter (see especially section 2.3): for Kant, our *sui generis* human responsibility for our actions is essentially shaped by the fact that we are imperfectly rational agents who always have the 'double option' to respect or reject the spirit of objective (moral) laws of reason (i.e., who can always exercise their general power for reflective self-control in a good or in a defective manner). This, I think, is Kant's main reason for insisting that human responsibility (unlike divine responsibility) requires alternative possibilities.

This brings us to (2). The important issue here is how one should interpret the notion of *possibility* that is employed in the assertion that an agent could have acted differently. Many compatibilists claim that this notion can be interpreted in a metaphysically innocuous way that does not require the absence of causal necessitation. I have previously tried to show that compatibilist proposals of this type are unsatisfying (cf. chapter VI, section 2.1). The gist of this discussion, applied to the issue of responsibility, is as follows. The problem with the famous Hobbesian-Moorean conditional analysis of 'can' ($R \text{ could do } x = R \text{ would have done } x \text{ if } R \text{ had chosen to do } x$) is that it provokes the question of whether an agent could have satisfied the condition mentioned in the analysis, i.e., of whether R could have chosen to do x . Some compatibilists have responded to this difficulty by proposing to interpret 'can choose' in an epistemic sense (choices x, y, z are all possible for a deliberator *for all she can know*).³¹⁹ This proposal may be sufficient to defend the relevance of the epistemic sense of 'can choose' for the forward-looking stance of agency (I argued in chapter VI that it is not), but it cannot secure the relevance of the epistemic sense of 'can' for the backward-looking stance of responsibility: with regard to choices that *have occurred* there is no longer any epistemic indeterminacy. Other compatibilists have despaired of the conditional analysis and have placed their emphasis on a *general, unanalyzed* notion of a capacity; compatibilists of this type emphasize that human agents, unlike (say) dogs, have a *general ability* to recognize and respond to norms that is not threatened by determinism.³²⁰ The compatibilist appeal to the notion of a general power of rational self-determination at this stage of the dialectic – i.e., in response to (2) – must be sharply distinguished from Wallace's earlier appeal to this notion (in response to (1)). Wallace explicitly repudiates the idea that the notion of a general capacity can be invoked in defense of the idea that agents could have acted otherwise than they in fact did.³²¹ And indeed (as I argued at some length in chapter VI), an appeal to general abilities is too unspecific to delineate what an agent is free to

³¹⁹ Moore 1912, p. 94 and Bok 1998, p. 120.

³²⁰ Frankfurt 2003a; Pettit/Smith 2003; Reath 2006, pp. 255-6; Scanlon 2003; Watson 2003; Wolf 1990, 2003.

³²¹ See Wallace 1994, pp. 208-10.

do on a concrete occasion.

Thus, neither the fact that we would do something different if we made certain choices nor the fact that we have the general ability to make certain choices seems by itself sufficient to show that we are free to make those choices *under the assumption that* the exercise of our powers of choice is exclusively governed by natural laws. So premise (2) of Kant's incompatibilist argument seems defensible as well. However, one may object here that even if canonical compatibilist interpretations of 'possible' are unsatisfying, it is incumbent on anyone who puts forward a claim like (2) to provide *her own* interpretation of 'possible' that renders (2) correct. I agree, and I shall turn to Kant's conception(s) of possibility in the following section.

I want to conclude my discussion of compatibilist attacks on Kant's argument by noting a general problem concerning compatibilist conceptions of responsibility and possibility. The problem is that compatibilists have a hard time respecting what we can call the 'temporal symmetry principle'. This principle states that we think about responsibility and possibility in exactly the same manner regardless of whether we occupy a forward – or a backward-looking stance. That is: when I deliberate about what to do, I consider myself responsible for future choices I *might* make in exactly the same manner as I consider myself responsible for past choices I *have* made; and this is because I consider alternative future choices *possible* in exactly the same sense in which I consider it possible, in hindsight, that I could have made alternative choices. Suppose I deliberate about whether I should lie to my partner to avoid an unpleasant conversation; suppose I choose to do so, and as a result my partner gets into trouble. When I deliberate about whether to tell the lie, I acknowledge that it is up to me to lie or not to lie, and that I shall *therefore* be blameworthy for the lie and for the bad consequences that may result from it. When my partner, having found out about the lie, confronts me with those consequences, I (ought to) acknowledge, in hindsight, that it was up to me to lie or not to lie and that *therefore* I now have to take blame for those consequences. My backward-looking thoughts *correspond to* my previous forward-looking ideas. Neither my conception of the possibility of different choices nor my sense of my responsibility for those choices is susceptible to a shift across time.

These seem to me to be perfectly common-sensical, pre-theoretical thoughts. And yet it seems endemic to compatibilism that it cannot accommodate them.³²² First, those compatibilists who suggest that the sense in which we consider various choices 'possible' while deliberating about what to do is exclusively *epistemic* (I can choose x, y, z for all I can know)³²³ must admit that this is not the sense in which we think, *in hindsight*, that various choices were possible: for there is no longer any epistemic indeterminacy in the stance that looks back at choices that have already occurred. (I can say, looking back, that I could have acted otherwise for all I knew *at that time*. But for this to be sense in which I declare, in hindsight, past choices that I did not make 'possible', I would have to abstract from my present state of knowledge of what I chose and instead simulate the mental state of ignorance that I was in before I made the choice; and this interpretation of our stance toward our past acts strikes me as utterly artificial. Moreover, we use the notion of epistemic possibility *only* because employing this notion is a necessary condition for occupying the forward-looking stance: I could not deliberate about whether to do x, y, z if I knew that I was going to do x. When I look back at my past acts, I no longer deliberate, and

³²² A possible exception might be Lewis-style compatibilism; but this kind of compatibilism introduces technical distinctions of its own that have no obvious repercussion in ordinary thought. See footnote 275 above.

³²³ See Bok 1998; Korsgaard 1996b; Moore 1912; Wallace 1994.

hence there is no point in using the notion of epistemic possibility from this perspective.) Second, those compatibilists who reject the idea that responsibility requires alternative possibilities cannot take at face value the apparent fact that when I deliberate about what to do, I acknowledge that I will have to take the blame for bad consequences of a bad choice *because* I acknowledge that alternative choices are genuinely open to me. Those compatibilists must *either* say that my self-conception as a responsible agent here is mistaken *or* that this self-conception changes (i.e., no longer requires alternative possibilities) when I look back at past actions. Third, those compatibilists who hold that our thinking about responsibility requires only an appeal to our general power for rational self-governance have to dismiss or reinterpret the idea that for a deliberating agent, as well as for someone who looks back at a past act, the idea that she has, or had, the general power to perform an act is consumed by the idea that she can, or could, exercise this power (in more than one way) in concrete acts of choice.

The fact that compatibilism violates the temporal symmetry principle does, I think, give voice to the incompatibilist worry that compatibilism imposes certain distortions upon our self-conception as responsible agents.

2: A tension between Kant's libertarianism and incompatibilism

I have argued that Kant has a defensible argument for incompatibilism about responsibility that rests on two commitments: first, responsibility requires alternative possibilities; second, a universal determinism does not leave us with such possibilities. This, however, raises the question: what notion of 'possibility' is Kant operating with in the above argument? In particular, what notion of possibility is at issue in his contention that if the exercise of all our capacities were governed by deterministic laws, we could never act otherwise than we in fact do?

Kant has two official concepts of possibility: *logical* and *real* possibility. A concept is logically possible just in case it does not contain contradictory predicates in its content (CPR, A 221-2/B 268-9). The concept of a man with two heads is logically possible, whereas the concept of a round square not. But it seems that Kant cannot be working with a notion of logical possibility in premise (2) of the above argument: (2) would be false if 'could' were to be understood in terms of logical possibility. If all our causal capacities had to be understood deterministically, it would still be *conceivable* that agents could have done something other than what they were necessitated to do. Thus, it seems as if Kant's premise (2) requires his notion of real possibility. Kant distinguishes between the real possibility of pure and empirical concepts. A *pure* concept is really possible if it conforms to the formal conditions of experience³²⁴ and applies a priori to all possible objects of experience (see CPR, A 220/B 267). With respect to *empirical* concepts, Kant explains (cf. CPR, A 223/B 270) that their real possibility depends on whether the concept has a foundation in experience and its known laws. For example, the concept of the power of foresight is not really possible because the laws of nature provide no reason to suppose that such a power exists.

The question is whether the notion of real possibility can help us to interpret premise (2) in such a way that it comes out true. Suppose I defamed someone in order to gain a professional

³²⁴ That is, if its objects 'can' be given in space and time. The modality here cannot be logical possibility, because if it were one could infer the legitimacy of the categories (i.e., their a priori applicability to experience) from the trivial fact that there is no logical contradiction in supposing that they can be instantiated in the phenomenal world. Kant can say that a pure concept is really possible if it is really possible that it can be instantiated in space and time. The right-hand side of the conditional specifies a new clause, so there is no circularity.

advantage. The claim that Kant needs for his incompatibilist argument is this: if all my capacities had to be understood in terms of the deterministic causality of nature, we would have to conclude that it was not really possible for me not to defame that person. Now, the empirical concept of a *general capacity* to tell the truth is perfectly in accordance with the laws of nature. Experience clearly shows that people *in general can* tell the truth even when tempted to lie. So, if 'could' is understood in this manner, premise (2) of Kant's argument is rendered false. Hence, to establish his incompatibilism about responsibility, Kant needs a notion of 'real impossibility' of individual *actions*. That is, he must claim that if all causality were mere nature, it would be really impossible for any agent to act differently than she in fact did. Now, Kant *does* speak of the real possibility of "substances, forces, reciprocal actions" (CPR, A 223 B 270); but can we attribute to him the notion that *individual* events are (in his technical sense) *really* possible or impossible?³²⁵

I want to argue that we can. To support this, one could appeal to epistemic considerations³²⁶, but the main point here is a *modal* one. Kant's notion of empirical *necessity* applies to individual events ("everything which happens is hypothetically necessary") and is defined in terms of the laws of nature: it applies to the effects that result from the exercise of general capacities "in accordance with empirical laws of causality" (CPR, A 227-8/B 280-1). If some event x occurs at time t, it occurs necessarily in accordance with the laws of nature given certain antecedent conditions. Now, trivially, whatever is empirically *necessary* must also be empirically *possible*. Since (as we saw) *both* the technical notion of what is 'really possible' *and* the notion of what is empirically necessary are defined by reference to the laws of nature, we should say that the kind of empirical possibility that corresponds to the notion of empirical necessity of events is real possibility: if event x is empirically necessary at time t, it is really possible at t. Now, intuitively, if an event is empirically necessary at t, the non-occurrence of this event at t is empirically impossible.³²⁷ If empirical possibility corresponds to real possibility, then empirical impossibility must correspond to real impossibility. It follows that if event x is empirically necessary at t, the non-occurrence of x at t is really impossible.

Hence, if I tell a lie at time t, my act of telling the lie is empirically necessary, since it is governed by laws of nature and being so governed is the sole criterion of empirical necessity. *Not* telling the lie at t is empirically (or, in the technical sense, *really*) impossible for me precisely because it would require a violation of the laws of nature.³²⁸ So if I lied at time t, it was

³²⁵ Watkins 2004 claims that Kant thinks of the exercise of capacities as continuously exercised activities. If this interpretation were correct, it would be trivially true that we can apply the notion of real possibility to activities, because on this model the notion of a causal capacity collapses into its continuous exercise. But even if we concede that this interpretation might be adequate to Kant's conception of powers of fundamental physics, such as the force of universal gravitation, and hence with regard to the notion of *reciprocal (inter)action*, it makes little sense in the context of a discussion of capacities such as the 'power of foresight' (Kant's own example), or in the context of a discussion of capacities that we ascribe when we hold people responsible.

³²⁶ Kant thinks that the notion of action has epistemic priority over that of substance (CPR, A 204-6/B 250-1) and over the notion of power (CPR, A 207/B 252), because we can infer the existence of substances and powers only through observing their effects. It would be rather odd if we could deem substances and powers 'really possible' but had no notion of real possibility that we can apply to those very events that entitle us to the belief in the real possibility of substances and capacities in the first place.

³²⁷ Kant *is* prepared to say that if an event x occurs necessarily at time t, the non-occurrence of x at t is impossible. KpV, 5: 95: "When I say of a man who commits a theft that, by the law of causality, this deed is a necessary result of the determining causes in preceding time, then it was impossible that it could not have happened."

³²⁸ One might object that the action might instead be brought about by a change of the past. But it is not really possible for an agent to change what has already happened. Experience and its laws give us no reason to suppose

not really possible for me not to lie at *t*. This seems to be what Kant needs for premise (2) of his incompatibilist argument: if there were only the deterministic causality of nature, no agent *could*, in the sense of real or empirical possibility, have acted differently than she in fact did, and judgments of responsibility would be illegitimate.

From the fact the exercise of my general ability for telling the truth at *t*₁ is really impossible, it does not follow that I do not possess this general ability. The general capacity for telling the truth is *itself really possible* since there are empirical conditions (which differ from those that obtain at *t*₁) under which it is possible for agents to exercise this capacity (contrast: the power of foresight). There is only one modal standard here: we establish *both* the real possibility of general powers *and* the real possibility of their exercise in particular circumstances by asking what the laws of nature allow: a question that concerns *both* the general constitution of objects *and* the specific activity flowing from this constitution under specific empirical conditions.³²⁹

I have attributed to Kant an incompatibilist argument to the effect that if there is only the deterministic causality of nature, agents do not have alternative possibilities and judgments of responsibility are illegitimate. Now, Kant does *not* think that judgments of responsibility are illegitimate. So he must think that it *is* typically possible for agents to act differently. But what sense of possibility does he employ in his positive conception of responsibility? I pointed out earlier that Kant has only two concepts of possibility, real and logical possibility. Now Kant himself *cannot* deem the performance of an action that is ruled out by the laws of nature – such as my telling the truth at *t*₁ when I in fact lied at *t*₁ – really possible in the strict, technical sense. The capacity of a transcendently free Willkür that is required for choosing and acting "contrary to [the] force and influence [of natural causes]" (CPR, A534-5/B562-3) is itself not really but only logically possible, because the concept of a free Willkür is, for Kant, an idea of pure reason that violates the formal conditions of experience (it cannot be represented in time) and hence fails the criterion (specified above on p. 139) for the real possibility of pure concepts.

But then Kant seems to be facing a dilemma. If he accepts that judgments of responsibility imply that alternative actions are really possible, he can establish the incompatibilist conclusion that if there is only natural causality, judgments of responsibility are illegitimate. However – this is the first horn – Kant is then committed to hard determinism because Kant's appeal to the non-natural causality of reason does not entitle *him* to the notion that alternative actions are really possible. Faced with this difficulty, Kant might concede that judgments of responsibility require only that alternative actions that are attributed to the non-natural causality of freedom are *logically* possible. But – this is the second horn – in this case premise (2) of Kant's incompatibilist argument is falsified: his compatibilist, naturalist opponent is *also* entitled to the claim that the conception of a free Willkür and of an alternative action is logically possible.³³⁰ The dilemma here is that *either* Kant's bars for the legitimacy of judgments of responsibility are so high that his own positive account fails to pass them, *or* these bars are so low that the account of his naturalist-compatibilist opponent easily passes them.

One may respond that the dilemma vanishes if we acknowledge that judgments of

that it is in anyone's power to effect such a change. See KpV, 5: 96.

³²⁹ We attribute general capacities to types of objects that fall under sortal terms whose content is partly given in terms of causal powers. This allows us to distinguish among types of objects by appeal to their typical capacities; and this serves our classificatory purposes. Since the empirical conditions that allow for the exercise of a general ability at a specific time are typically repeatable, attributing such abilities is informative in a deterministic world.

³³⁰ Allison 1990, 2004 sometimes seems to deny this; I discuss and reject his interpretation in chapter II, section 3,

responsibility are made from a practical point of view. That is, it might be held that Kant thinks that, given the primacy of practical over theoretical reason (KpV, 5: 119 ff.), occupying the practical point of view allows us to bracket speculative questions about the metaphysics of possibility and to assume that it is always really possible *in a practical sense* for agents to do what duty requires.³³¹ But this suggestion cannot by itself absolve Kant from facing the above dilemma. The question is: does the appeal to the idea that judgments of responsibility are made from a practical perspective require that alternative actions must be 'really possible' in the strict sense of the term (that involves a reference to the conditions of experience and the laws of nature) in order for such judgments to be legitimate? If 'yes', we still get impaled on the first horn of the above dilemma. If the answer is 'no', and Kant's *only* point here is the primacy of practical reason that renders speculative issues about possibility idle, he must give up on his incompatibilist argument. For the primacy of practical concerns, and the idea that such concerns preempt metaphysical doubts about what is possible for us to do, can readily be invoked by someone committed to a naturalistic metaphysics. This confirms my general conviction that an appeal to practical concerns cannot be *all there is* to Kant's position: that is, such an appeal cannot altogether replace metaphysical patterns of justification.³³² If the appeal to the practical standpoint allowed us to sidestep metaphysical worries about the compatibility of determinism and responsibility, Kant's conception of responsibility would collapse either into a compatibilism that employs the metaphysically shallow conception of possibility that Kant so relentlessly repudiates (as a 'wretched subterfuge'), or into the view (held, on occasion, by Peter Strawson³³³) that our reasons for ascriptions of responsibility are dictated by pragmatic considerations that render all theoretical concerns about determinism irrelevant.

Can Kant escape this dilemma? We must realize here that Kant's incompatibilist argument is directed against a transcendental realist. This means, among other things, that Kant and his opponent will have a different understanding of the general point of modal talk. For Kant, modal talk about 'objects' (in a broad sense, including capacities and actions) has epistemic rather than ontological significance; it indicates the relation of representations to our epistemic faculties. To say that something is possible in the schematized sense, i.e., 'really possible', is to say that it accords with the conditions of experience or the laws of nature. To say that something is possible in an 'unschematized', purely formal sense expresses that its concept is thinkable (CPR, A 219-20/B 266-7). By contrast, for Kant's realist opponent, talk about modality amounts to an insight into the absolute ontological constitution of objects that entitles us to claims about what properties they could have and what actions they could perform absolutely, *in themselves*, independently of their relation to our epistemic faculties. Now, Kant's incompatibilist argument is *more specifically* directed against someone who has an exclusively empiricist or naturalistic worldview but still seeks to hold onto our self-conception as free, responsible agents.³³⁴ It is characteristic of this empiricist branch of realism that it construes what are, in Kant, mere

³³¹ One could appeal here to Kant's suggestion that the notion of practical freedom has a kind of 'objective reality', and can thus be regarded as 'really possible', in a "merely practical" sense. See KpV, 5: 28-30.

³³² See my chapter II, and footnote 305 above.

³³³ See P. Strawson 2003, pp. 82-3.

³³⁴ This may not be obvious, for in the Transcendental Dialectic, where Kant is most explicitly concerned with the empiricist/naturalist branch of transcendental realism, he represents the empiricist as a hard determinist who denies that we are free. But in the Critique of Practical Reason, Kant attributes the compatibilist conception of freedom that he goes on to reject specifically to the empiricist qua universal determinist (KpV, 5: 94).

conditions of objects of our experience as conditions of objects in general³³⁵: conditions such as, 'actions must occur in time and in accordance with necessitating empirical laws'.

Kant and his naturalist opponent *agree* that an appeal to time, to the general deterministic character of natural causality, and to specific natural laws imposes constraints on which capacities and actions are possible in a thick, empirical or 'real' sense that goes much beyond the constraints imposed on possibility by formal conditions of logical coherence. This agreement explains why both Kant and the naturalist deem a capacity for absolutely free choice, whose exercise is not constrained by conditions of time and of natural causality, impossible *in some sense*. But the decisive difference is that for Kant, this verdict means *only* that the possibility of a free Willkür violates epistemic conditions of experience and (hence) that we cannot theoretically know (for instance, on the basis of intuition or perception) that we are free. By contrast, the empiricist (qua transcendental realist) thinks that a free Willkür violates ontological conditions of being (conditions such as time or the deterministic character of all real causality); and this renders freedom from causal necessitation metaphysically impossible in an absolute sense.

Now, for both Kant and the naturalist, everything that is empirically or really possible is also logically possible, while some things are *only* logically possible. For the naturalist, to say that something is not empirically but *only logically* possible is to imply that it is merely thinkable but has no ontological foundations: its conception is 'empty'.³³⁶ By contrast, when Kant says that something is merely logically possible, he means not that it is contrary to an absolute ontological condition that something has to satisfy in order to come to be but that it cannot be represented as part of the course of experience. The phrase, 'merely thinkable' has disparaging *epistemic* connotations for Kant and disparaging *ontological* connotations for the naturalist.

What do these considerations show with respect to the above dilemma? The dilemma I expounded is that if Kant interprets 'could' in his incompatibilist argument in terms of 'real possibility', then premise (2) is true but Kant is committed to hard determinism; and if he interprets 'could' in terms of 'logical possibility', then (2) is falsified and the incompatibilist argument fails. This dilemma arises under the assumption that Kant and his compatibilist opponent interpret modal talk in the same manner: if they did, then it would be unfair for Kant to commit his opponent to the idea that judgments of responsibility require that alternative actions are really possible and to ground his positive account on the idea that judgments of responsibility require only that alternative actions are logically possible. But given their different interpretations of modal talk, pursuing this strategy is not unfair. Judgments of responsibility imply that it was possible for an agent to act differently in a *relevant* sense, where this means *at least* that the representation of the possibility to act otherwise must not conflict with ontological principles that constrain what can come to be. For Kant's naturalist opponent, whatever is merely thinkable but not really possible is contrary to principles that constrain what can come to be, and hence *for her* it is irrelevant, regarding judgments of responsibility, whether an alternative action was merely logically possible. However, *for Kant*, it need not be irrelevant, as far as judgments of responsibility are concerned, that alternative actions are merely logically possible, because for him the notion that an action is *merely* thinkable means *only* that the action is contrary to principles that constrain everything we can represent as part of our framework of objective

³³⁵ See CPR, A 781/B 809 for the distinction between two kinds of transcendental realism.

³³⁶ Hence the empiricist's claim, in the antithesis to the Third Antinomy, that the concept of an unconditioned causality (the idea of transcendental freedom) is "an empty thought entity" (CPR, A 447/B 475).

experience, *not* that it is contrary to principles that constrain everything that can come to be.

In short: Kant and his naturalist opponent agree that the laws of nature constrain the exercise of all our empirical capacities. But the naturalist affirms, whereas Kant does not affirm, that our empirical capacities exhaust our metaphysical constitution and our causal powers.

In the light of these considerations, we can represent Kant's incompatibilist argument in the following, refined way: (1) To judge that someone is responsible for her action is to imply that she could, *in an ontologically relevant sense*, have performed a different action than the one she in fact performed. (2) If there were only the deterministic causality of nature (i.e., if (d) were true), no agent could, in this relevant sense, have acted differently than she in fact did: for in this case empirical or 'real' possibility (in the strict sense, as defined by reference to the laws of nature) would be the *only* ontologically relevant modal category, and this category rules out, as we saw, the possibility of actions that did not in fact occur. (3) If (d) were true, judgments of responsibility would be illegitimate because they imply something that is strictly false.

To clarify: Kant holds that our acts, considered as natural (phenomenal) events, *are* causally necessitated; but he holds that our acts can *also* be attributed to a noumenal cause whose causality operates outside the deterministic order. This is a corollary of the point that for Kant empirical (im)possibility is not the only modally relevant category because he does not accept that our ontological constitution is exhausted by our empirical capacities. The idea that our phenomenal acts and choices have both a (causally necessitated) phenomenal or empirical ground and a (free) noumenal ground presupposes Kant's idealist framework: in this framework, space, time and deterministic causality do not provide ultimate ontological constraints on what is and can come to be, but are (merely) conditions imposed on objects and acts by our epistemic faculties. Objects are spatiotemporally situated and causally determined not in themselves but as appearances (in relation to our faculties of cognition).³³⁷ We can attribute our acts to both a free noumenal cause and to a causally determined empirical cause only because we can accept (given Kant's idealism) that an appeal to empirical causality does not disclose the ultimate ontological layer of reality. It does disclose an ultimate epistemic layer – the bounds of empirical causality are, for beings like us, the bounds of theoretical knowledge. Hence we cannot (as I indicated in section 1) expect the appeal to the noumenal grounds of acts to provide us with a framework within which our free, responsible agency can be made theoretically comprehensible. This is the price that Kant's conception of free, responsible agency requires us to pay.

I take myself to have shown that Kant has an argument against compatibilism about moral responsibility that is consistent with the softness of his incompatibilism. But a caveat is in order here. My argument *only* shows that Kant's idealism allows him to hold, against a naturalist who thinks that the laws of nature impose ultimate constraints on what can come to be, that there is metaphysical space for accommodating the ontological requirements of judgments of responsibility. I have *not* shown that Kant establishes that this space is indeed filled, i.e., that we *have* a free Willkür that allows us to act in a way that is not (as far as the noumenal grounds of our actions are concerned) constrained by the deterministic laws of nature. It is for the purposes of showing, against the background of Kant's idealist framework, *that we really are free* that the appeal to the practical standpoint and in particular to the consciousness of the moral law can play an argumentative role, as giving us *positive reasons* for adopting the belief in the kind of

³³⁷ See chapters I and II for elaboration on these ideas.

freedom required for responsibility that Kant's idealism (merely) makes room for.³³⁸ I examine what argumentative weight the appeal to the moral law can play in chapter XIII.

3: The primacy of inner agency in Kant's theory of responsibility

According to my reconstruction of Kant's account of responsibility, he thinks that the legitimacy of judgments of responsibility can be saved since his idealism makes room for the notion that our capacities are not exhausted by our empirical powers, and hence for the idea that alternative actions are possible in a metaphysically relevant sense. Now, one might worry here that this appeal to empirically unconstrained capacities cannot give us a viable conception of responsibility. If we remove the bounds of empirical possibility, don't we open up the floodgates for assigning responsibility for all kinds of acts that are intuitively beyond an agent's power?³³⁹ For instance, it seems that we need an appeal to what is physically possible to explain why a man who is locked in iron chairs is not to be blamed for not saving a drowning child.

I suggest that Kant should respond here by pointing out that his account does not give us reasons to believe in *just about any types* of non-empirical capacities. As I intimated above, his idealism *only makes room for* the idea that not all our capacities are constrained, in their exercise, by natural laws and empirical conditions. Positive reasons for belief in such capacities require an additional source. Although, as I said, discussing Kant's justification of the belief in freedom is beyond the scope of this chapter, we must here acknowledge that it is only our awareness of the moral law that positively justifies our belief in the non-empirical capacities that are required for responsibility. Since our awareness of our power to act in ways that are unconstrained by the laws of nature derives specifically from our awareness of being subject to moral requirements, we can only assume as much freedom from nature as is required for complying with moral laws. Now, *what* moral norms require is the formation of volitional structures, that is, choices of maxims (principles of willing) that subordinate our contingent desires to the value of humanity as an end in itself. Hence, our awareness that we actually can comply with moral requirements entitles us *only* to the assumption that it was possible to *will and choose* independently of natural conditions (for instance, that someone who made a bad choice could have chosen otherwise). Our awareness of our freedom to comply with moral norms *does not legitimize* a belief in 'freedom of body'. Correspondingly, *what we are morally responsible for*, on Kant's conception, are (primarily) inner acts of choice rather than outer, bodily acts. Thus, Kant's account of responsibility requires and allows alternative possibilities, and hence requires and allows that agents are unconstrained in their activity by empirical laws, only at the 'inner' level of volition.³⁴⁰ It does not license the idea that we can exercise our

³³⁸ See chapter II for an elaboration and defense of these ideas.

³³⁹ This problem might be considered a version of Bennett's complaint (1974, p. 222) that "Kant's theory will not let us draw any usable line around the actions for which the agent is accountable" and implies an 'indiscriminately universal' responsibility. It must be noted that this worry by no means afflicts only Kant's theory. It is a challenge for any (broadly libertarian) account of responsibility that holds that our capacities are not ultimately constrained by natural laws to non-arbitrarily distinguish between (a) actions that intuitively lie within the scope of responsibility, and that are to be deemed possible because the agent is not constrained by the laws of nature; and (b) actions such as breaking iron chains, that intuitively fall outside the scope of responsibility, but which seem to require no more than what actions falling under (a) require, namely, that we can break the laws of nature.

³⁴⁰ One could recast the worry at the inner level and hold that Kant's account implies that agents are responsible for their willing in cases where we would intuitively judge that it was not really possible for them to *will or choose differently*. Such cases might include, first, scenarios where the agent is mentally impaired or insane. Bennett

physical capacities in ways that are unconstrained by empirical conditions and forces.

This conception of moral responsibility rests on the idea that complying with or violating a moral law, and hence being deserving of praise or blame, is a matter of respecting or failing to respect the force of moral reasons; and responding to or failing to respond to the reasons represented by moral 'oughts' is primarily a matter of how one chooses. That is to say: *the act* that essentially reflects an agent's respect or disrespect for the force of moral reasons is an inner act of choice, rather than an outer act that involves the exercise of bodily capacities. Whether or not a choice results in the intended good or bad effect at the level of outer behavior depends on empirical conditions that are at least partly beyond the agent's control. By contrast, making a good or bad choice is always in every rational person's control. Whether or not a person complies with the moral law and is thus to be regarded as praise – or blameworthy can relate only to what is always fully within her control; hence, agents are to be held responsible primarily for their choice of a general moral disposition (*Gesinnung*), as contained in maxims.³⁴¹

1984 claims that Kant must concede that an insane person is responsible for her acts, but Kant can say that such a person lacks (like animals) a spontaneous faculty of practical reason that would enable her to form maxims; hence, she cannot be considered an imputable agent. A second type of case is that where agent choose under conditions of duress (stress, anxiety, etc.). Here we can acknowledge that it is *harder* for someone under such conditions to will in ways required by the moral law, and this can mitigate our judgment of responsibility; but this is quite consistent with saying that it was possible for her to resist the influence of those factors, and hence that she is responsible for her volition. Third, an agent might be under the influence of drugs or hypnosis. Here Kant could say that the power to form maxims in response to moral reasons is temporarily blocked: the kind of non-empirical capacity that allows one, in normal circumstances, to choose independently of causal necessitation is, under these special circumstances, no longer available; the agent is temporarily removed from the space of reasons. One might object that it is slippery to hold that a capacity to choose independently of empirical necessity can be blocked by empirical conditions. Here Kant must rely on the idea that there is a grave difference between the constraints that empirical laws and conditions puts on normative reflection and choice in ordinary circumstances and the constraints exercised by such conditions under special circumstances such as hypnosis or drugs. In a way, Kant is no worse off here than his compatibilist opponent, who *also* has to suppose that it is only in these special circumstances that empirical laws and conditions impugn our capacity for rational control. Moreover, Kant and the compatibilist are in the same boat in another sense: neither of them can *give an account of how* our ability to respond to normative reasons is blocked by (say) drugs. There is no causal, descriptive story that may shed light on why our capacity for *normative* reflection is disabled by conditions such as drugs, without giving away the normative phenomenon. The naturalistically inclined compatibilist, like the libertarian, must rely here on the brute idea that such conditions temporarily push us outside the space of reasons.

³⁴¹ KpV, 5: 36-7: "It is always in everyone's power to satisfy the categorical command of morality; whereas it is seldom possible, and by no means so to everyone, to satisfy the empirically conditioned precept of happiness... The reason is that in the former case there is question only of the maxim, which must be genuine and pure; but in the latter case there is question also of one's capacity and physical power to realize a desired object." KpV, 5: 45: "We are here concerned only with the determination of the will and the determining principles of its maxims as a free will, not at all with the result. For, provided only that the will conforms to the law of pure reason, then let its power in execution be what it may...". GMS, 4: 394: "A good will is good not because of what it performs or effects, not by its aptness for the attainment of some proposed end, but simply by virtue of the volition; that is, it is good in itself, and considered by itself is to be esteemed much higher than all that can be brought about by it (...). Even if it should happen that, owing to special disfavor of fortune, or the niggardly provision of a step-motherly nature, this will should wholly lack power to accomplish its purpose, if with its greatest efforts it should yet achieve nothing, and there should remain only the good will (not, to be sure, a mere wish, but the summoning of all means in our power), then, like a jewel, it would still shine by its own light, as a thing which has its whole value in itself. Its usefulness...can neither add nor take away anything from this value." GMS, 4: 435: "... fidelity to promises, benevolence from principle (not from instinct), have an intrinsic worth...their worth consists not in the effects which spring from them...but in the disposition of mind, that is, the maxims of

It is important to be clear here about the dialectical structure of Kant's position. Kant does *not* assign responsibility to inner acts of choice because he is worried that his libertarian idea that our responsible acts are (as far as their noumenal grounds are concerned) unconstrained by empirical laws would have implausible implications if it were applied to bodily acts. Rather, he is driven to his assignment of responsibility to inner acts of choice on the basis of the idea that our choices to respect or disrespect moral 'oughts' (which are the true and proper object of judgments of moral responsibility) are only imperfectly correlated with our outer, observable behavior; and this idea does *not* depend on incompatibilist intuitions. (After all, as we saw, sophisticated compatibilists such as Jay Wallace *also* hold that what we are responsible for is the quality of our will.) We can see this by considering the following, uncontrived example. Suppose that Sue is firmly committed to keeping her promise to pick up her daughter when school ends. Now suppose that due to a fire incident, school ends two hours earlier, and that Sue has no way of knowing. Here empirical conditions beyond Sue's control prevent her from translating her morally good volition into a corresponding effect at the level of outer behavior; and yet she deserves moral praise, not blame. This example supports the idea that inner acts of choice or volition rather than instances of outer behavior are the primary focus of judgments of responsibility. But the example does not make use of any point that a compatibilist must find objectionable, and it would support Kant's assignment of moral responsibility to inner acts *even if* we assumed that agents had the kind of 'freedom of body' that Kant's theory disallows. Even if agent's were unconstrained, in their bodily movements, by the laws of physics, there would remain unforeseeable vagaries of the empirical world which render the correlation between one's response to moral norms and one's outer behavior imperfect and accidental.

The notion that we are responsible primarily for our volitional activity and that compliance with the moral law requires only the right choices stands in need of clarification. This notion does *not* imply that moral laws do not give us concrete prescriptions about what is to be done in the spatial, observable world. Rather, the moral primacy of the inner (volitional) level means that any prescription to bring about a certain observable effect is, *at bottom*, a prescription to respect moral values and hence to form a volition that subordinates everything to these values. Such willing *implies* that in certain empirical circumstances one strives with all one's power to bring about certain observable effects. Suppose I come across a shallow pond and see a child struggling against drowning. Here the prescription to treat humanity as an end in itself requires that I strive, with all my physical power, to save the child. This shows that the primacy of inner activity pertains only to contexts of assessment, namely, contexts where we raise questions about the moral quality of an agent's will and about whether she is deserving of praise or blame. The inner level does *not* have primacy for an agent who seeks to determine what to do. The fact that what we happen to cause at the outer level is only partly within our control is consistent with the fact that our volitional striving is directed toward causing physical effects. From the agent's point of view, intending the good is equivalent to doing all in her physical power to (say) help those in need or to do what was promised. The moral law impinges on an agent's consciousness in the form of a prescription to do all one can to respect the unconditional value that it represents. *What it demands is not success but volitional commitment to success*; but this difference is deliberately irrelevant for a morally good agent.

the will which are ready to manifest themselves in such actions, even though they should not have the desired effect."

One might object here that we cannot coherently picture agents as intending to bring about physical effects through exercising their bodily capacities if we attribute to them the belief that they are not free to exercise their bodily capacities as they intend to. In response, I would insist that while one cannot intend to perform a bodily act 'b' if one knows that performing b lies not in one's physical power, intending to perform a bodily act b is perfectly compatible with believing that it is possible but to some extent uncertain whether one shall succeed in performing b and that whether or not one shall succeed in performing b is at least in part beyond one's control. Indeed, these beliefs accurately describe our deliberative self-conception: we intend to perform bodily outer acts that we believe (with greater or lesser confidence) to be within our general physical powers, but we can never rule out that these powers may leave us in the lurch (i.e., that we may be unable to exercise them in the manner we intend on a specific occasion).

Kant is never perfectly clear about these issues, but I think that the following remark confirms that he acknowledges the kind of deliberative self-conception I just sketched:

The action to which the 'ought' applies must indeed be possible under natural conditions. These conditions, however, do not play any part in determining the will itself, but only in determining the effect and its consequences in the [field of] appearance. (CPR, A 548/B 567)

Kant here distinguishes between an act that must be possible under natural conditions and the act of determining the will (for which natural conditions are deemed irrelevant); furthermore, the act that must be possible under natural conditions is considered an effect of willing in the field of appearance. Undoubtedly, this act is what Kant calls an outer act that unfolds in space, that involves the use of bodily capacities, and that can be considered an observable appearance. In what sense of 'possibility' must such an action be represented as 'possible' by an 'ought'? Going back to what I argued above in section 2, I suggest that what Kant has in mind here is the notion of empirical, real possibility as applied to general bodily capacities rather to concrete exercises of such capacities. That is: "possible under natural conditions" here means roughly, 'possible given the way we are, generally speaking, bodily constituted'. The moral law prescribes that we strive with all that generally lies in our physical power to do certain things. Whether or not it will be actually empirically (really) possible to exercise the relevant physical powers on a concrete occasion is, typically, both unknown to us and never fully within our control, and hence we are primarily responsible for our volitions. For example: respecting the value of humanity, as prescribed by the moral law, requires for any ordinarily constituted agent that she tries to rescue a child who is struggling against drowning in a shallow pond, because it lies within the general physical powers of most human agents to do this. It may, however, turn out that exercising this power in the concrete case is made impossible by empirical conditions: one may slip, or find oneself paralyzed, or find one's arms too weak, or the pond's current may be too strong, etc.; in short: given the state of the physical world (including one's body), the laws of nature may prevent one from physically executing one's volition to pull a person out of a pond by blocking the exercise of one's general physical power to do so. What the above passage brings out nicely is that Kant holds that the moral law presupposes that the exercise of our *powers of choice* is *not* likewise constrained by empirical conditions: the determination of our will that leads to a morally good choice is an absolutely free inner act which is always within a rational agent's power to perform. We are, at least typically, free (noumenally, not phenomenally) to choose in accordance with or contrary to moral reasons independently of causal necessitation.

Kant's insistence on the moral primacy of the inner (volitional) level does not imply that

inner acts of volition cannot cause physical effects. Kant clearly acknowledges that we think that our free choice of maxims *may* have such effects (cf. CPR, A 547-9/B 575-6). However, Kant thinks that our physical behavior is only a secondary and indirect effect of the exercise of our freedom of will (whose primary and immediate effect is the formation of a certain volitional structure). That we will to do one thing rather than another is, in cases of intentional activity, a necessary condition of the occurrence of a particular physical effect but not sufficient for bringing about this effect. The impact of our volitional, inner activity on our physical, outer behavior is mediated by the interference of factors that obtain independently of what we will, and hence the physical effects of our maxims are, *unlike* the formation of the maxims themselves, not *fully* within our control. Moreover, even when the outer effects of an inner principle are fully intended, there is an element of what I might call 'moral contingency': to what extent a fully intended outer action reveals the moral character of the agent is always dependent on empirical conditions and circumstances that are beyond the agent's control. A genuinely good-willing agent may not be presented with opportunities to prove her moral worth just as a genuinely malicious agent may not be presented with opportunities to prove her moral deviance.

For Kant, these considerations show that outer actions are not the true focus of judgments of responsibility. However, this is compatible with holding that outer actions play an indispensable role in our practice of assigning responsibility: volitional states are unobservable, and hence only outer actions provide us with (inconclusive) warrant for the ascription of unobservable inner states that we consider to be the true seat of responsibility.³⁴² So while an agent is strictly responsible only for her choices, we can infer the moral quality of those choices only by observing the way in which they 'manifest' themselves at the 'outer' level.³⁴³ This epistemic significance of outer actions accounts for the fact that we actually blame A, who performs a morally bad outer action, rather than B who behaves in a morally neutral manner even though she also has adopted a bad maxim: we may simply lack grounds for the attribution of a bad maxim to B. If an unchained man refuses to save a drowning child, his moral badness is manifested by his outer act; in the case of a chained man who has the same bad maxim, we may lack grounds for the attribution of a mental state that calls for blame.³⁴⁴ Hence, we can say that for Kant the significance of outer actions for judgments of responsibility is that they are (potentially misleading) *symptoms* of moral character.³⁴⁵

Let us go through different types of cases to see the precise implications of Kant's theory of moral responsibility. First, there are cases where an outer action is a *fully intended* effect of willing: if the principle of willing is dutiful, the agent is praiseworthy both for her outer action and for the maxim which brings about her observable behavior (even though we may never be fully warranted in attributing to an agent a good maxim on the basis of her observable behavior);

³⁴² GMS, 4: 407: "...we can never, even by the strictest examination, get completely behind the secret springs of action; since, when the question is of moral worth, it is not with the actions which we see that we are concerned, but with those inward principles of them which we do not see." Compare RGV, 6: 20.

³⁴³ See RGV, 6:20: "We call a man evil, however, not because he performs [outer] actions that are evil...but because these actions are of such a nature that we may infer from them the presence in him of evil maxims."

³⁴⁴ I think that one could make a case for the claim that the observation of actual deeds is much more conclusive, for ascribing maxims to an agent, than the agent's reporting to us what her maxims are. This is not only because of concerns about the genuineness of the report but also because people are very prone to moral self-deception. Kant thinks that no agent can know what her own moral character is.

³⁴⁵ This may also be part of the point of Kant's notoriously obscure remark that a person's empirical character, as displayed in her observable acts, is the "sensible sign" of her noumenal character (CPR, A 546/B 574).

if her willing is contrary to the moral law, the agent is blameworthy both for her outer action and for the maxim which brings about her observable behavior (even though we may never be fully warranted in deeming the agent evil, as opposed to weak-willed or 'frail'³⁴⁶).

Second, there are cases where a genuinely good principle of willing fails to bring about a good observable effect. Suppose that I am moved by my awareness of the unconditional value of humanity to try to save a child from drowning; but I slip and fall unconscious. Here, I am praiseworthy for my willing, although (of course) not for my outer action (and there may be no warrant for attributing a good maxim to me). The same diagnosis holds where a genuinely good willing has bad effects through no fault of the agent: for example, when a man rushes to help a person who is attacked by criminals, slips, and thereby causes a bystander to run against a car.³⁴⁷

Third, there are cases where a genuinely bad principle of willing does not manifest itself in bad outer actions. These are the cases which show that it is a matter of 'moral contingency' to what extent the physical world provides us with opportunities to reveal a bad moral character. Consider a merchant A who has adopted the maxim, 'I will cheat on my customers whenever I face no risks', and who happens to be visited only by very powerful people whom it would be dangerous to cheat on. Compare A to a merchant B who has adopted the same maxim, is visited by less powerful people, and cheats on them. Merchant B exhibits the first type of case: she is to be blamed both for her (fully intended) bad outer action and for her bad maxim. But A is equally *deserving* of blame, despite the difference in outer behavior.³⁴⁸ The reason why A is equally deserving of blame as B, even though his outer actions are permissible, is that we should not give A any credit for being unable to prove his moral unworthiness. It is a matter of moral luck for A not to run into customers that he deems safe to exploit; but A does not deserve moral credit (that would be implied by the judgment that he is not blameworthy) just because there are empirical conditions beyond his control that prevent his morally bad maxim from manifesting itself at the level of outer activity. Of course, to the extent that there is no observable symptom of moral corruption, we may lack evidence for attributing the bad maxim to A.

Another case that falls into this third category further supports the point that it is irrelevant for a judgment of responsibility whether a person has alternative possibilities *at the level of outer activity*. Suppose C has promised to spend the evening with a friend, D, who is in great distress. On his way to D's place, C decides to rather go bar hopping. Now suppose that C could not have gone to D's place anyway because there has been an accident so that the only road to D's place is blocked. Here C is blameworthy even though the outer act of keeping his

³⁴⁶ An outer action that is contrary to duty may betray either the agent's evil will or her frailty (RGV, 6: 29). See RGV, 6: 20 for the claim that the inference from an observable action to an *evil* maxim is never fully justified.

³⁴⁷ Of course, an agent who has a genuinely good principle of willing cannot be oblivious to these consequences: she will try to make up for them in some way. The idea that the physical results of willing may be irrelevant to our judgments of responsibility does not imply that they are also irrelevant for the question what the agent morally ought to do. This point is rightly emphasized by Herman 1996, pp. 97-8.

³⁴⁸ 'Treating someone as a mere means' is a matter of adopting an evil inner principle that shape one's interactions with others, even if this 'shape' is not observable: the merchant who does not cheat on a customer solely because the customer is capable of harming the merchant nonetheless treats the customer as a mere means; he does not refrain from cheating on the customer out of respect for the customer, but solely as the result of a calculation of the possible effects of interactions with the customer. It is an example of this type that Kant has in mind when he says (RGV, 6:31): "The maxim, then, in terms of whose goodness all moral worth of the individual must be appraised, is thus contrary to the law, and the man, despite all his good deeds, is nevertheless evil."

promise was impossible for him to perform.³⁴⁹ Likewise, in our earlier example, if the chained man who is forced to watch a child drown has adopted a maxim of indifference that would dispose him not to save the child even if he were unchained, this man is deserving of blame despite the fact that he lacks alternative possibilities at the level of outer behavior.

In the above scenario, merchant A performs morally permissible outer actions (i.e., acts 'in accordance with duty') on the basis of an impermissible, bad maxim: A is ready and willing to cheat provided only that he thinks that he can get away with it. We can distinguish here another merchant C, who is firmly disposed to treat his customers honestly all the time, but for purely egoistic reasons (because it is to his advantage, and he is very risk-averse; see GMS, 4: 397). It is clear that C is not deserving of moral praise for his acts. But is C *blameworthy*? I think that the answer must be 'yes'. There are two ways of defending the intuition that it would be arbitrary to hold that A is blameworthy whereas C is not. First, C displays the same kind of disrespect toward others as A, and he is equally oblivious to the weight of moral reasons and values: the only difference between A and C is that C is more risk-averse, but this is a contingent fact about his psychology that should bear no moral relevance. Second, all it would take for C to actually cheat on his customers is a change in his empirical circumstances: he might wake up tomorrow and feel more inclined to take risks, or his risk-averseness might become increased by a drastic decrease of his financial resources. This shows that C's performing morally permissible or even required outer acts is not an expression of his moral character but a function of how his egoism is constrained by the vagaries of the empirical world.³⁵⁰ Hence, C is deserving of blame, although we may lack evidence for attributing a morally bad maxim to him.

A more tricky kind of case is that in which an agent D perform a morally required outer act on the basis of what Kant calls (GMS, 4: 397-8) 'immediate inclination'. For instance, suppose D sees someone lose one thousand dollars and returns the money on the basis of her warm feelings for others, that is, on the basis of her benevolent temperament that makes it hard for her to bear the distress of others and that makes her rejoice over their happiness. In chapter VI, section 3, I defended Kant's idea that an agent who acts on the basis of such an inclination is not deserving of moral praise. In a nutshell, the idea is that the objective 'ought' of morality provides necessary, desire-independent reasons for acting, and thus a choice based on a contingent desire (such as the want to make others happy) does not respect the normative force of moral 'oughts'. Here the outer act (such as returning the money) contains *the letter but not the spirit* of the objective norm (KpV, 5: 71-2). Now, even if one concedes that an agent who chooses on the basis of contingent benevolent desires is not deserving of moral praise, the tricky question is whether the agent is deserving of moral blame. Kant himself is not perfectly

³⁴⁹ This example is a variation of one given by Wallace 1994, pp. 141-2.

³⁵⁰ I am not claiming that *any* counterfactual specification of a scenario in which an agent would perform a morally bad outer act suffices to show that an agent is undeserving of moral praise or even deserving of moral blame. If an agent acts from duty, the fact that he would violate moral norms if he was faced with the same choice tomorrow (say, because his moral strength has waned, or because he has meanwhile adopted a bad maxim) does not, on my view, show that he does not deserve praise for his actual deed. Thus, my appeal to counterfactuals is not subject to the criticism that Herman 1996, p. 10 makes of the appeal to counterfactuals in determining moral worth (I agree with her that we should distinguish moral worth from virtue). The systematic difference here is this: in imagining that someone who actually acts from duty makes a bad choice, we are stipulating a change of her moral character; but in imagining that someone who is actually oblivious to the weight of moral reasons performs (due to a mere change in empirical circumstances) morally impermissible outer acts, we are keeping her moral character fixed. We thereby make vivid the inappropriateness of her *actual* motives .

consistent in his answer to this question, but I think that his considered and justified answer is, or ought to be, 'yes'.³⁵¹ The reasons are the same as in the two previous cases. First, someone who helps others on the basis of her contingent affectionate temperament is oblivious to the force of moral reasons. Second, if D's general practical orientation (that is, her maxim) is founded upon her affectionate want to see others happy, then her outer actions are – just like those of merchants A and C – morally permissible only by accident: this maxim will lead to morally impermissible outer acts depending on the circumstances (see chapter VI, section 3 for some examples). Hence, there is a perfectly clear sense in which D's motives can be subjected to moral criticism, which is to say that D is deserving of blame for discarding the force of moral reasons and for assigning weight to contingent desires that lead to morally permissible or required outer acts only by accident. Again, it may be very hard, even impossible to determine whether or not D acts on the basis of desire-dependent or desire-independent reasons, and so our best bet may be to give her the benefit of the doubt and to praise her for what she does (the reason being that actions based on contingent benevolent desires are, *at least in the world as it is*, more likely to lead to outer acts that are in accordance with duty than actions based on egoistic desires).

It does not follow that there are no choices or outer acts that are neither morally praise – nor blameworthy. We can posit a fourth category for morally permissible choices made for the sake of realizing contingent desires. My choice to buy a particular brand of yogurt because I find its taste delicious may fall into that category. This does not mean that moral considerations are altogether irrelevant to choices of this type. My choice to buy yogurt cannot be morally praiseworthy (since it is done for desire-dependent reasons, rather than for the characteristic desire-independent reasons represented by moral 'oughts'), but it *might* very well be morally blameworthy: this might be the case if the maxims underlying my nutritional choices or my spending habits are oblivious to considerations of health or global justice. It is only if my non-moral willing is (however implicitly) informed by my sensitivity to its permissibility that my specific choices are neither morally praise – nor blameworthy.

An altogether different, fifth category concerns the bad consequences resulting from bad maxims. Kant holds that the agent is to be blamed for his bad choice and *all* bad effects that may result from it.³⁵² Now, the bad effects resulting from a bad choice may include (i) intended bad consequences; (ii) unintended bad consequences; and (iii) consequences that could not have been foreseen by the agent. If condition (i) is satisfied, Kant's position is obviously compelling. Kant's view also seems compelling in cases where only condition (ii) is satisfied: if my bad choice results in unintended bad effects that I could have foreseen if I had been more careful or thoughtful, the fact that I did not intend those consequences cannot absolve me from moral criticism – my lack of care and thought is incorporated into the criticism. Kant's view seems problematic, however, in a case where both (ii) and (iii) are satisfied. We saw that a good agent is

³⁵¹ In the Groundwork (GMS, 4: 398), Kant holds that an action based on a desire to make others happy deserves "praise and encouragement". He clearly does not have 'moral praise' in mind here. I suggest that his point is (as I go on to say in the main text) that such desires are, in the world as it is, more likely than egoistic desires to lead to actions that *at least accord with duty*. But Kant is committed to holding that *any* case where an agent makes her contingent desires the grounds of her choice calls, regardless of the object of those desires, for moral criticism and blame. In the Religion, his claim that "The maxim, then, in terms of whose goodness all moral worth of the individual must be appraised, is thus contrary to the law, and the man, despite all his good deeds, is nevertheless evil", is explicitly meant to apply to maxims based on 'benevolent instinct'. See RGV, 6:30.

³⁵² MM, 6:228: "...the bad consequences of a wrongful action...may be imputed to the agent."

not responsible for bad effects that result through no fault of her own (i.e., that she did not intend and could not foresee); how, then, could it be consistent to impute the unintended and unforeseeable bad effects of bad maxims to the agent? One obvious difference between the two cases is that if unforeseeable bad consequences result from a good choice, the agent made no discernible mistake at all; whereas the bad unforeseeable consequences resulting from a bad choice can be traced to an avoidable moral fault. The reason why this matters is as follows. The excuse, 'I did not want this to happen to you, and I could not have foreseen it!' is wholly effective only if it can be combined with the truthful addition, 'and I respected you as a person'. For if I know that my wrongdoer acted out of disrespect for me and my dignity (which, as is worth noting, is quite compatible with the fact that she wished me no harm), then I know that her claim that she did not *want* those unforeseeable things to happen to me, even if true, does not alter the fact that she treated me as a mere means or instrument, and thus subordinated my dignity to the satisfaction of her private desires. This means that the 'want' that her excuse refers to is itself a morally inappropriate attitude, which deprives the excuse of its moral force. Notice here that Kant's position can tolerate the idea that the fact that the bad consequences of a bad choice were both unintended and unforeseeable by the agent *mitigates* the blame that the agent deserves for those consequences. Kant's position requires only that the agent is not completely absolved from facing moral criticism for those consequences. And this point does seem true to the phenomena: suppose I tell a lie to avoid an unpleasant conversation with my partner, and as a result my partner gets into trouble – something I did not intend and (we can suppose) I could not have foreseen. Here it does seem entirely appropriate, even compelling, for my partner to blame me by saying: 'You should have thought twice before making your personal satisfaction the condition of your choice.' Hence, there is something to be said for Kant's point that by willing the bad and disrespecting the weight of moral reasons, we forfeit our right to be wholly excused for outer effects that result from the unforeseeable vagaries of the empirical world.³⁵³

In sum, Kant's conception of moral responsibility is focused on the idea that human agents are responsible for their 'inner agency', namely, for their choices of general maxims and of specific (outer) acts. It is always within a rational agent's control to choose for or against the reasons represented by moral 'oughts'. We are blameworthy if we fail to choose for those reasons. We are praiseworthy if we succeed in making morally good choices, that is, if we succeed in resisting the influence of contingent desires that impel us to choose on their behalf.³⁵⁴

4: Moral versus juridical responsibility

I want to conclude by noting that the conception of responsibility sketched thus far relates only to *moral* responsibility. Kant thinks that in contexts where we determine the *legal* consequences of actions, a different conception of responsibility is at issue, whose standards are considerably less demanding than those required by judgments of moral responsibility. What matters here is not someone's volitional structure, but merely whether the outer activity of people is in accordance with what the juridical law requires. Consider again our two merchants A and B, who have adopted the maxim, 'I will cheat on my customers whenever I occur no risk'. Suppose B is caught cheating whereas A refrains from cheating since his customers all happen to be very

³⁵³ For a somewhat different way of justifying Kant's point, which I find quite congenial, see Reath 2006, pp. 260-3.

³⁵⁴ For my defense of the idea that human agents are always tempted to choose contrary to moral norms, and for a qualified defense of the idea that Kant holds a 'battle citation' model of moral worth, see footnote 283 above.

powerful. We saw that according to the standards imposed by what Kant calls 'ethical legislation' (i.e., the standards of *moral* responsibility), A and B are equally deserving of blame. According to the standards of what Kant calls juridical legislation and responsibility, B is to be held fully responsible *for her outer action* whereas A is *not to be held responsible at all*, because here we are only concerned with the legality of the action and not at all with the motive that leads the agent to keep the law.³⁵⁵ A second difference is implied by the following case: suppose J and K are paid killers, and J (due to a superior physical constitution) succeeds whereas K fails in killing the victim. According to the standards imposed by ethical legislation, J and K are equally blameworthy. But according to the standards of juridical legislation, K is to be punished less harshly than J, because attempted murder is less severe a crime than murder.

Hence, standards of juridical responsibility do not always put weight on the fact that the outer effects that result from an agent's choice are contingent (which is why they give credit to someone who, due to physical circumstances beyond her control, fails to fully execute her bad intentions), and they are not concerned with the purity of the agent's motive (which is why they exempt someone who performs legal acts on the basis of fear of consequences, greed, etc.). It seems to me that this difference between the standards of ethical and juridical responsibility must go along with a difference between the concepts of the possibility of alternative actions that is employed by the two standards. The notion of mere logical possibility of alternative actions is suitable only in a context that is primarily concerned with the purity or impurity of inner, volitional activity, and where what happens at the external level has the status of a mere symptom. The juridical context with its focus on the legality of what happens at the external level requires a notion of possibility that tracks mainly the absence of external coercion.

I suggest that it would make sense for Kant to allow that for the purposes of assigning *legal responsibility*, or for settling whether an agent is deserving of *punishment*, it may be enough to determine whether she *would have acted differently had she chosen to do so*; for if this question is to be answered negatively, it follows that it was not in her physical power to omit her outer action, in which case she is not legally responsible for it (whereas this does not at all settle questions about *moral* responsibility). The conditional notion of 'can' that tracks the freedom of outer activity from external coercion seems ideally suited in this context. The question, 'could she have chosen differently?' that the conditional analysis of 'can' naturally invites is not irrelevant here, but jurisdiction can (in Kant's system) assume that the answer is 'yes' in an unconditional sense that signals complete control over volitional activity. While the conditional sense of freedom of 'outer' action answers to the purposes of jurisdiction, this pragmatic context does not float free of worries about the compatibility of determinism and responsibility: it is *especially* acts of punishment that would seem profoundly *unfair* if agents were, in virtue of being causally determined at the level of volition, not responsible for their crimes.³⁵⁶ The non-

³⁵⁵ MM, 6: 218-9: "All legislation, therefore, may be differentiated by reference to its motive-principle. The legislation which makes an action a duty, and this duty at the same time a motive, is ethical. That legislation which does not include the motive-principle in the law, and consequently admits another motive than the idea of duty itself, is juridical. In respect of the latter, it is evident that the motives distinct from the idea of duty, to which it may refer, must be drawn from the subjective (pathological) influences of inclination and of aversion, determining the voluntary activity, and especially from the latter; because it is a legislation which has to be compulsory, and not merely a mode of attracting or persuading. The agreement or non-agreement of an action with the law, without reference to its motive, is its legality; and that character of the action in which the idea of duty arising from the law at the same time forms the motive of the action, is its morality."

³⁵⁶ See Wallace 1994, pp. 103-9 for the idea that considerations of fairness are integral to incompatibilists.

pragmatic, morally relevant notion of the possibility of an unconditionally free choice can be called upon to answer those worries and to terminate the regress that threatens the conditional analysis when it is applied within a purely deterministic framework.³⁵⁷

³⁵⁷ Bojanowski 2006, pp. 44-5 also defends the claim that the absolute concept of freedom is important to Kant's conception of legal responsibility, but for reasons exactly opposite to mine. He claims that Kant would (i) be happy to justify punishment along compatibilist-utilitarian lines (by appeal to the effects on future behavior) but would (ii) insist that legal requirements aspire to determine the will categorically and purely in a way that rules out motivation through sensuous causes. I think that nothing can be further removed from Kant's views on punishment than (i) (see KpV, 5: 37) and the above passage from the *Metaphysics of Morals* (cf. footnote 355) shows conclusively that juridical legislation does not require purity of motivation.

Part 3
**From (theoretical) freedom of thought to (practical)
freedom of will**

Chapter VIII: The Normative and the Empirical Standpoint

In the preceding chapters, I expounded the grounds of Kant's incompatibilism about practical freedom. We still face the question of what could possibly justify us in thinking that we really *are* practically free. It is sometimes held that the key move in Kant's attempt to legitimize the idea of freedom is his distinction between standpoints. Interpreters who put a lot of weight on Kant's appeal to standpoints tend to argue that this appeal is meant to supplant a metaphysical account of our self-conception as free agents.³⁵⁸ This, if correct, would effectively remove the need to legitimize the idea that we are free in a metaphysically thick sense. However, I have argued (see, especially, chapters II, V, and VI) that Kant's conception of freedom cannot be metaphysically deflated. Thus it is incumbent on me to show what Kant's standpoint distinction is meant to contribute to his justification of the idea of freedom. In this chapter I provide a detailed interpretation of Kant's conception of standpoints; on this basis, I try to understand what work his standpoint distinction is doing in his theory of freedom. It will turn out that an adequate interpretation of his standpoint distinction brings into focus the issue of *theoretical* freedom.

1: Allison's standpoint distinctions

The seminal interpretive work of Henry Allison has shaped many discussions of Kant's appeal to standpoints, so I shall begin by reviewing the heavy use that Allison makes of the notion of a standpoint. One peculiar aspect of Allison's account is that he never quite analyzes what Kant's own standpoint distinction is supposed to be. Rather, he introduces a variety of standpoint distinctions that are only loosely (if at all) based on what Kant says, but that are nonetheless supposed to do some systematic interpretive work. There is (1) the contrast between the 'standpoint of transcendental idealism' and the 'standpoint of transcendental realism'.³⁵⁹ This contrast tracks different 'metaphilosophical' commitments (e.g., that we can or cannot have knowledge of things in themselves).³⁶⁰ Second (2), the rejection of "the assumption that there must be some standpoint-independent fact of the matter"³⁶¹ is supposed to be relevant to the question of whether we are free or causally determined: our freedom "is only assertible from a 'practical point of view'", whereas causal determinism is true from a theoretical point of view.³⁶² Third (3), there is the distinction between the empirical standpoint (concerned with "first-order investigation of objects") and the transcendental standpoint (concerned with "second-order philosophical considerations of objects and the conditions of their cognition").³⁶³

Whether or not Allison intends it, these three distinctions are irreducibly distinct from one another. (1) cannot be identified with (2) or (3) because the standpoint of transcendental realism is defective, whereas the standpoints that figure in (2) and (3) are perfectly legitimate. (2) and (3) cannot be identified either: while the theoretical standpoint in (2) presumably corresponds to the empirical standpoint in (3), the transcendental standpoint from which one conducts "second-order philosophical considerations of objects and the conditions of their cognition" in (3) is not the practical standpoint of (2), from which *cognition of objects* is a non-issue. Indeed, Allison's

³⁵⁸ See Allison 2004, pp. 47-8 and Korsgaard 1996a, p. x (in her 'Introduction').

³⁵⁹ Allison 2004, pp. 22-23.

³⁶⁰ Allison 2004, p. 28; p. 35; pp. 394-95; p. 445: the two standpoints are "metaphilosophical alternatives".

³⁶¹ Allison 2004, p. 47.

³⁶² Allison 2004, p. 48.

³⁶³ Allison 2004, p. 63.

discussion renders the very concept of a standpoint ambiguous: his talk about standpoints is either (as in (1)) meant to refer to specific (meta)philosophical commitments or (as in (2) and (3)) meant to track the difference between a range of (practical, theoretical, first-order, second-order) questions. These are very different ways of understanding what a standpoint is. Now, I do not see that Kant himself ever uses the notion of a standpoint in the context of elucidating the 'metaphilosophical' commitments of transcendental realism or idealism. So we should disassociate Kant's notion of a standpoint from what Allison intends with (1). It remains to be seen whether (2) or (3) capture the essence of Kant's standpoint distinction; I shall argue that neither (2) nor (3) *quite* does.

2: Kant's standpoints as perspectives of representation

There are only two places in Kant's critical writings where a distinction between two different standpoints is explicitly invoked: in the Transcendental Aesthetic of the Critique of Pure Reason and in the third section of the Groundwork. Since I believe that Kant's discussion in the Aesthetic is, despite its extreme vagueness, essential for understanding his talk about standpoints on the whole, I shall first examine this discussion.

In the relevant passage from the Aesthetic (CPR, A 26/B 42-A 27/B 42) Kant mentions a "human standpoint". It is clear that he thinks that from this standpoint, we believe that space and time are *real* and contain objects that have spatiotemporal properties, such as shape, size, or duration.³⁶⁴ Now it might seem as if the human standpoint is, as a matter of tautology, the *one* standpoint that human beings can occupy, so that any other standpoint must be a *non-human* perspective that is beyond our reach. If that were correct, it would follow that in the Aesthetic Kant cannot be concerned with drawing a contrast between two different standpoints that we can occupy. This is affirmed by Allison, who claims that "the 'human standpoint'...is the only standpoint available to us"³⁶⁵ and by Beatrice Longuenesse, who thinks that the human standpoint is to be characterized as a *standpoint of discursive judgment* that requires both receptive and spontaneous faculties: our spontaneous 'capacity to judge' can be exercised only under conditions of sensible affection, and this contrasts with the perspective of animals (that are merely receptive) and the perspective of God (who is merely spontaneous).³⁶⁶

The characterization of a perspective on the world in terms of spontaneity and receptivity is indeed sufficient to distinguish the human perspective from the animal and from the divine perspective. But Kant in the Aesthetic wants to distinguish the human standpoint from the perspective "of other thinking beings" (CPR, A 27/B 43), and he points out that we are in no position to ascertain whether the intuitive representations of these other thinking beings are subject to the same *conditions of sensibility* as ours. This shows that the perspective of other thinking beings is meant to be a perspective of non-human discursive thinkers, not the perspective of God (whose thinking is not subject to any conditions of sensibility) or animals (who do not think). So the appeal to the discursive nature of human thought is not sufficient to give us Kant's notion of a human standpoint, because it does not capture the specifically human element in the way we represent objects and the world. Given *our human* forms of sensibility,

³⁶⁴ For analogous remarks about time and temporal representation, see CPR, B A 33-37/B 49-53. The analogy is not perfect: time does not afford us with representations of enduring objects (CPR, B 155-6 and B 276).

³⁶⁵ Allison 2004, p. 132.

³⁶⁶ Longuenesse 2009, pp. 2-3. She thinks that the interplay of receptivity and spontaneity is characteristic of the human capacity to judge also in moral and aesthetic contexts.

our discursive thought is of a *spatiotemporal* world of sense. But it is at least conceivable that there are beings that have different forms of sensibility and whose discursive thought is not of a spatiotemporal world. Such beings would be both receptive and spontaneous, but would not occupy the human standpoint. Hence, the human standpoint must essentially be a *spatiotemporal* perspective on the way things are.

This, by itself, does not rule out the idea that the human standpoint is the only standpoint available to human beings. If the human standpoint essentially involves the notion of characteristically human forms of sensibility, then the question is whether human beings might occupy a distinctive perspective without relying on their forms of sensibility. More precisely, the question is whether we might be capable of a *kind of representation* of objects that does not depend on our specifically human faculties of sensibility. Kant intimates that this is possible when he appeals (CPR, A 28/B 44) to objects as "they are considered in themselves through reason, that is, without regard to the constitution of our sensibility." This suggests that human beings are capable of 'considering' or representing objects in terms of concepts *of reason* that do not depend on the 'constitution of their sensibility'. These concepts cannot be cognitively determinate representations that allow for theoretical knowledge of objects. For such representations, human beings depend on receiving data through sensible faculties whose forms are space and time: "What objects may be in themselves, and apart from all this receptivity of our sensibility, remains completely unknown to us" (CPR, A 42/B 59). But for Kant the conditions of knowledge are not identical to conditions of thought. Employing the ('unschematized') categories of the understanding independently of their relation to our sensibility allows us to *think* certain formal features (CPR, A 147/B 186-7; A 243/B 301); and these forms of thought can acquire object-directedness from the ideas of reason. For instance, when the notion of a ground that has something else as its consequence (the unschematized category of causality) is applied to the idea of spontaneity or freedom, it relates to an unconditioned causality. This representation is, because of its lack of perceptual or intuitive content, theoretically indeterminate: the idea of freedom is fruitless for the purposes of observing, explaining, understanding, and predicting objects and their behavior.³⁶⁷ And, (again) because the content of the unschematized categories abstracts from sensible data and the transcendental idea of freedom likewise "contains nothing borrowed from experience" (CPR, A 533/B 561), the representation of free, spontaneous causality could be shared by other, indeed by

³⁶⁷ KpV, 5: 136: "Every employment of reason in respect of an object requires pure concepts of the understanding (categories), without which no object can be conceived. These can be applied to the theoretical employment of reason, i.e., to that kind of knowledge, only in case an intuition (which is always sensible) is taken as a basis, and therefore merely in order to conceive by means of them an object of possible experience. Now here what have to be thought by means of the categories in order to be known are ideas of reason, which cannot be given in any experience. Only we are not here concerned with the theoretical knowledge of the objects of these ideas, but only with this, whether they have objects at all. This reality is supplied by pure practical reason, and theoretical reason has nothing further to do in this but to think those objects by means of categories. (...) Now when the categories are to be applied to these ideas, it is not possible to give them any object in intuition; but that such an object actually exists, and consequently that the category as a mere form of thought is here not empty but has significance, this is sufficiently assured them by an object which practical reason presents beyond doubt...". Through the pure, unschematized categories we think an object in general, but this thought is altogether 'empty' or indeterminate (CPR, B 150; A 399); the representation achieved by the pure categories is rendered determinate and acquires a relation to an object through the ideas of pure reason, although this object-directedness is still insufficient for the theoretical use of reason. See my chapter I, especially section 5 for further discussion.

all thinking beings even if they lack human sensibility and are incapable of human experience.³⁶⁸

So there is no paradox involved in supposing that there is a standpoint from which human beings can conceive objects in abstraction from specifically human forms of sensibility. This standpoint can be regarded as a 'more than human' standpoint: the representations we come up with from this standpoint are representations that derive from faculties that we (would) share with other thinking beings, namely, reason and the understanding. It contrasts with the *merely* human standpoint, from which we come up with representations that depend on our merely human forms of sensibility. In sum, the standpoint distinction in the Aesthetic suggests that Kant's standpoint distinction tracks two irreducibly different ways of representing objects.

3: Kant's standpoints as perspectives of reasoning

But the appeal to different ways of representing objects cannot be all there is to Kant's standpoint distinction: the suggestion that we can distinguish between two standpoints in terms of the distinctive representational resources available to each standpoint leaves open the question what the point and purpose of employing those resources is supposed to be. What do we seek to achieve when we represent objects in terms of the two types of concepts (the 'merely human' and the 'more than human' ones) that Kant associates with his two standpoints? Kant's discussion in the Aesthetic does not answer these questions. We must look to the Groundwork here.

It must first be acknowledged that the standpoint distinction in the Groundwork is exactly the same as that in the Aesthetic:

One resource remains to us, namely, to inquire whether we do not occupy different standpoints when by means of freedom we think ourselves as causes efficient a priori, and when we form our conception of ourselves from our actions as effects which we see before our eyes. (...) ...we must admit and assume behind the appearance something else that is not an appearance, namely, the things in themselves; although we must admit that as they can never be known to us except as they affect us, we can come no nearer to them, nor can we ever know what they are in themselves. This must furnish a distinction, however crude, between a world of sense and the world of understanding, of which the former may be different according to the difference of the sensibility in various observers, while the second which is its

³⁶⁸ One might worry here that an abstraction from our forms of sensibility does *not yet* give us the concept of things in themselves, because it suppresses the distinction between (I) the concept of an object of a non-sensible intuition and (II) the concept of an object of sensible but non-human intuition; only (I), it might be held, gives us the concept of a thing in itself. Here we need to take into account an important distinction that we first encountered in chapter I. Kant's *negative* concept of a noumenon is that of a thing insofar it is not an object of *our* sensible intuition; the *positive* concept of a noumenon is that of an object of non-sensible, intellectual intuition. We *have* a conception of things in themselves in the negative sense, namely, a mere negative correlate of our conception of a thing that has spatiotemporal predicates: *all* this concept requires are negations of predicates that we do possess (see CPR, B 149-50). By contrast, in the positive concept we think a non-sensible form of intuition; since we cannot understand the possibility of such intuition we do not have a cognitively determinate sense of the positive determinations that an object of such intuition would have (CPR, B 307 ff.). When Kant says that we can 'consider things in themselves through reason, without regard to the constitution of our sensibility', what he means is that the ideas of reason can give content (albeit not a content suited for the theoretical use of reason) to the *positive* concept of a noumenon: by employing the ideas of reason we get a positive conception of properties that, since they relate to the 'unconditioned', cannot be an object of *any* discursive form of experience (CPR, A 510/B 538) and hence must be considered 'intelligible' objects of a non-sensible, intellectual intuition. The 'neglected alternative' of the concept (II) of things of *sensible but non-human intuition* mentioned by the above worry satisfies the definition of Kant's a negative concept of a noumenon, i.e., the notion of something insofar it is not an object of our intuition. We have as little a determinate or positive conception of an object of non-human sensible intuition as we have of an object of intellectual intuition.

basis always remains the same. (GMS, 4: 450-1)

Kant holds that those of our representations that are subject to the conditions imposed by our forms of sensibility give us knowledge of a world of appearances that "may be different according to the difference of the sensibility in various observers". This parallels the Aesthetic's appeal to 'other thinking beings' that may not be bound by the conditions of human sensibility. Kant then emphasizes that we may also conceive objects independently of their relation to our (or indeed any) sensible faculties, as things in themselves; precisely because no reference is made here to the forms of sensibility that distinguish one kind of (discursively) thinking being from another, the representation of this 'intelligible world' is the same for all such beings. This parallels the Aesthetic's appeal to a standpoint from which objects "are considered in themselves through reason, that is, without regard to the constitution of our sensibility."

Now, in the Groundwork, Kant goes on to add what can be regarded as a specification of the purpose that is served, respectively, by occupying the two standpoints:

For this reason a rational being must regard himself qua intelligence (not from the side of his lower faculties) as belonging not to the world of sense, but to that of understanding; hence he has two standpoints from which he can regard himself, and cognize laws of the exercise of his faculties, and consequently of all his actions: first, so far as he belongs to the world of sense, he finds himself subject to laws of nature (heteronomy); secondly, as belonging to the intelligible world, under laws which being independent of nature have their foundation not in experience but in reason alone. (GMS, 4: 452)

Kant here indicates that by considering themselves from the two standpoints, rational beings can cognize two distinct types of *laws of actions*. Thus, from each of these perspectives, we are capable of a conception of ourselves and our activity in accordance with a distinctive set of laws. From the standpoint that gives us a representation of ourselves as spatiotemporally located, sensible objects, the relevant laws are causal laws of nature. From the standpoint that gives us a representation of ourselves as 'intelligible' objects, the laws we are concerned with are 'independent of nature'; these non-natural laws must be *normative laws*.

A standard interpretation of Kant's standpoint distinction holds that the standpoint (i) from which we represent ourselves as belonging to the sensible world is a theoretical or empirical standpoint from which try to explain and to predict our actions, whereas the standpoint (ii) from which we represent ourselves as non-sensible objects is a standpoint of practical deliberation from which we seek to determine our will.³⁶⁹ I accept one half of this interpretation: I agree that the standpoint (i) is an *empirical* perspective from which we attempt to represent a unified course and system of nature in which events (including our actions) can be observed, explained, predicted and systematically correlated by appeal to natural, causal laws. But the conception of standpoint (ii) that the standard interpretation proposes seems to me too narrow.

According to the standard interpretation, standpoint (ii) involves the exercise of our faculty of will or practical reason. But Kant characterizes (ii) by saying that from this perspective we cognize non-natural, non-causal laws "for the exercise of our faculties", thus referring to *faculties* in the plural. A close reading of this passage confirms the significance of his use of the plural: the first sentence makes clear that the faculties that Kant has in mind here are those that a rational being attributes to herself 'qua intelligence' and that are to be contrasted with the being's 'lower faculties'. By 'lower faculties', Kant means faculties with respect to which we are passive

³⁶⁹ This is Allison's (2), as described above. See also Korsgaard 1996a, Matthews 1969, Nelkin 2000, Watkins 2004.

and receptive, and whose exercise cannot be said to stand under normative laws.³⁷⁰ Hence, it is not the exercise of those lower faculties that we have in view when we cognize normative laws for the use of our faculties from the non-empirical standpoint (ii). What, then, are the *higher* faculties that characterize us as 'an intelligence' and whose exercise *is* governed by norms? The Transcendental Analytic (CPR, A 131) tells us: "General logic is constructed upon a ground plan which exactly coincides with the division of the higher faculties of knowledge. These are: understanding, judgment, and reason." So Kant's idea that from the non-empirical standpoint we recognize normative laws for the use of our higher faculties refers not merely to the faculty of *practical* reason, but also to theoretical reason, the understanding, and the power of judgment.

This suggestion can be strengthened. Before he draws the distinction between the two standpoints by appeal to the difference between causal and normative laws, Kant says :

Thus in respect to mere perception and receptivity of sensations [man] must reckon himself as belonging to the world of sense; but in respect of whatever there may be of **pure activity** in him (that which reaches consciousness immediately and not through affecting the senses), he must reckon himself as belonging to the intelligible world... (...) Now man really finds in himself a faculty by which he distinguishes himself from everything else...and that is reason. This being **pure spontaneity** is even elevated above the **understanding**. **For although the latter is self-activity** and does not, like sense, merely contain intuitions that arise when we are affected by things (and are therefore passive), yet it cannot **produce from its activity** any other concepts than those which merely serve to bring the intuitions of sense under rules and, thereby, to unite them in one consciousness...whereas, on the contrary, reason shows **so pure a spontaneity** in the case of what I call ideas that it thereby far transcends everything that the sensibility can give it.... (GMS, 4: 452; my emphasis)

It is clear from this passage that for Kant the non-empirical standpoint is one from which we engage in *purely spontaneous activity* that produces (accordingly) *pure* conceptual representations. It is on the basis of our capacity for such purely spontaneous activity that we consider ourselves 'as intelligences' that belong to the intelligible world. The crucial point here is that all of the aforementioned higher faculties, including the understanding, are capable of such purely spontaneous activity. In the above passage, Kant clearly means to refer to reason as such, *including* reason in its theoretical employment, as purely spontaneous: the 'ideas' he mentions include the ideas of theoretical reason. I want to argue for the more controversial point that Kant ascribes the same generic, pure type of spontaneity also to the *understanding*. This is denied, for instance, by Michael Friedman:

But the freedom required by morality should not be confused with the 'spontaneity' characterizing the understanding. Indeed, the difference between these two marks precisely the distinction between the faculty of understanding and the faculty of reason by Kant. The understanding, which necessarily operates in coordination with sensibility, secures the foundations for the modern idea of nature by

³⁷⁰ Kant's point here is actually quite difficult. From the text, it is clear that his notion of 'lower faculties', that relates to our membership in the sensible world, is meant to refer to "mere perception and the receptivity of sensation". But it should be clear that our forms of sensibility, and the passive (transcendental) affection of these forms by 'outer' objects, is not itself an aspect of the sensible world that stands under empirical laws of nature. I conjecture that Kant's point here is this. If we merely focus on the fact that a being has faculties of sensibility that allow for the passive reception of sensible data, we are not yet discriminating between rational beings and higher animals. While those animals may – by virtue of the fact that they have a non-sensible constitution in themselves – be said to belong to the intelligible world, they lack higher faculties of knowledge that allow for the *spontaneous* combination of data and whose exercise affords them with a *consciousness* of their membership in an intelligible world. Their combination of sensible data stands exclusively under empirical, psychological laws.

injecting the a priori rationality of mathematics and logic into sensible experience. The faculty of reason, by contrast, can operate entirely independently of sensibility, and its product – the moral law – is thus wholly free of all sensible contamination.³⁷¹

For our purposes, the central issue here is an ambiguity in Friedman's notion of an 'operation that is entirely independently of sensibility'. For this notion may refer, first, to the operation that *produces* a certain representation or, second, to the operation that *applies* or makes a use of this representation. Friedman is right to point out that there is a difference between the understanding and (both practical and theoretical) reason with respect to the second kind of operation. The categories (pure concepts of the understanding), unlike the ideas of reason, have (as we already saw) a determinate object only in relation to sensibility; hence, the categories are limited in their determinate application to what is given through sensibility. But this point pertains only to the sphere of application of the categories. It does not pertain to their *production*: when we focus on their production or origin, it seems clear that the categories of the understanding, *just like* the ideas of reason, are produced 'entirely independently of sensibility', which means that both these products are 'free of all sensible contamination' or, in Kant's terminology, that they both have a 'pure' origin.³⁷² Of course, the production of the categories is *occasioned by* impressions of sensibility (CPR, A 86/B 118), but the ideas of pure practical reason are also occasioned by desires that impel us to incorporate them, as grounds of choice, into our maxims.

This point can be confirmed by appeal to a passage from the Dialectic (CPR, A 546-9/B 574-77) where Kant draws a distinction between two ways in which human beings can consider themselves: as possessing capacities "the causality of which must stand under empirical laws", and as possessing faculties whose exercise is not conditioned by sensible conditions:

[man] is thus to himself, on the one hand phenomenon, and on the other hand, in respect of certain faculties the action of which cannot be ascribed to the receptivity of sensibility, a purely intelligible object. We entitle these faculties understanding and reason.

This is the same contrast between two distinctive perspectives as in the Groundwork. And, as in the Groundwork, the faculties whose exercise cannot be considered as standing under empirical laws or as being conditioned by sensible causes are the 'higher faculties' of understanding and reason, sources of pure representations. Again, Kant indicates that the understanding is subordinate to reason (in the sense that its concepts depend for their applicability on sensible data) without denying that the understanding produces "its own...pure concepts".

Hence, Kant associates the non-empirical standpoint that attests to our membership in the intelligible world with the purely spontaneous exercise of all our higher faculties, not merely with the purely spontaneous exercise of practical reason. Now, we saw that from this standpoint we cognize 'non-natural', i.e., normative laws of our actions. If we combine these points, the following picture emerges: from the non-empirical standpoint we engage in purely spontaneous activity whose products are certain concepts or ideas that contain normative laws for the use of our higher faculties. The purely spontaneous exercise of practical reason allows us to cognize

³⁷¹ Friedman 1996, p. 438. Compare Allison's claim (1990, p. 222) that Kant in the above passage draws "a sharp distinction between the spontaneity of understanding and that of reason", and the idea, advanced by Kitcher 1984, p. 122 and Sellars 1970, p. 26 that the spontaneity of reason is absolute whereas the spontaneity of the understanding is merely 'relative'.

³⁷² CPR, B 104-5: "Pure synthesis...gives us the pure concept of the understanding". See CPR, A 94/B 126-8 for Kant's insistence that the *origin* of the categories must be independent of sensibility.

laws for practical agency, and the purely spontaneous exercise of our higher epistemic faculties, such as theoretical reason and the understanding, allows us to cognize laws for epistemic agency.

A defender of the orthodox conception of the non-empirical standpoint as a merely practical point of view, such as Henry Allison, might object here that my suggestion that from the non-empirical standpoint we cognize normative laws that govern our epistemic agency is confused because the very notion of rational, norm-governed agency is confined to the exercise of our will.³⁷³ This objection raises an important question: does Kant have a substantive notion of rational, norm-governed *agency* that does not refer to the will? I shall argue that he has such a notion in chapters IX and X. For present purposes, we can notice at least that Kant liberally applies the notion of action to the exercise of our epistemic faculties; he says, for instance, that "thinking is the action by which given intuitions are related to an object".³⁷⁴

My idea that both the understanding and practical reason are considered by Kant as purely spontaneous sources of normative laws for agency is confirmed by the following passage:

What I call applied logic...is a representation of the understanding and of the rules of its necessary employment *in concreto*, that is, under the accidental subjective conditions which may hinder or help its application, and which are all given only empirically. (...) Pure general logic stands to it in the same relation as pure ethics, which contains only the necessary moral laws of a free will as such, stands to the doctrine of the virtues [i.e.] the doctrine which considers these laws under the limitations of the feelings, inclinations, and passions to which men are more or less subject. Such a doctrine can never furnish a true and demonstrated science, because, like applied logic, it depends on empirical and psychological principles. (CPR, A 54-5/B 78-9)

Notice that according to what Kant says here, the purity of both of pure general logic and pure ethics consists in their independence from empirical conditions: hence, the rules established by pure logic and pure ethics are, to use Friedman's phrase, free of all sensible contamination. Notice, further, that the laws of pure general logic are established through "actions of pure thinking" ("Handlungen des reinen Denkens"; CPR, A 58/B 82-3), just as the laws of pure ethics are established by actions of pure practical reasoning (GMS, 4: 448; CPR, A 803/B 831). Hence, it seems that the "rules of the necessary employment of the understanding" are analogous to the 'rules of the necessary employment of the will': both qualify as normative laws for (respectively) the exercise of our higher faculty of understanding or our higher faculty of will.

According to my interpretation of Kant's standpoint distinction, this distinction concerns two different manners of considering the use of our higher faculties: as standing under causal laws of nature (from the empirical standpoint) and as standing under normative laws of reason (from the normative standpoint). As we saw, the *power of judgment* is one of our 'higher faculties'. I now want to support my interpretation by noting that Kant draws precisely the distinction envisaged by my account, between considering our faculty of judgment empirically (psychologically), as it is actually exercised, and normatively, as it ought to be exercised.

If we propose to set forth the origin of these fundamental propositions [the principles of the power of judgment] and try to do so by the psychological method, we violate their sense. For they do not tell us what happens, i.e. by what rule our cognitive powers actually operate, and how we judge, but how we ought to judge.... (KU, 5: 182). Aesthetic judgments of reflection...do not state that everyone judges in

³⁷³ Cf. Allison 1990, pp. 221-2: "...in regarding ourselves as rational agents, that is, as rational beings with wills." Unless we had free will, we would be "rational beings but not rational agents; our sense of agency would be illusory" (*ibid.*, p. 63; compare p. 228).

³⁷⁴ CPR, A 247-8/B 304-5. See also A 68/B 93; A 77/B 102-3; A 80/B 106; B 130-1; B 143-4; B 154.

this manner – if they did, they would be a task for the explanations of empirical psychology – but rather that one ought to judge in this manner.... (EEUK, 20: 239)

I take this to be a particular application of the general contrast between the empirical and the normative perspective on our actions, here applied to actions such as *judging*. It is the same contrast that Kant also draws in contexts when he is specifically concerned with practical prescriptions. Hence, I suggest that Kant's standpoint distinction relates not a contrast between theoretical and practical reasoning but to a contrast between *empirical* and *normative* reasoning. This distinction is intended to track the fundamental difference between considering an agent with aim of observing, explaining, and predicting what she did, does or will do, and considering her with the aim of specifying what she should do or should have done. If we consider the agent from the standpoint of empirical cognition, we aim to observe, explain and predict her behavior by appeal to causal laws. If we consider the agent from the normative standpoint, we aim at specifying how she should act or have acted. And this general contrast can be drawn *both* with regard to our practical agency *and* with regard to our thinking and judging about nature.

4: Kant's standpoints as perspectives of representation and of reasoning

In section 2, I claimed that Kant's standpoint distinction is framed in terms of two distinctive kinds of representational resources that we can use to 'consider' objects; in section 3, I claimed that his standpoint distinction is framed in terms of two different aims that we seek to achieve when we occupy the two perspectives: the cognition of a normative order (i.e., of what ought to happen) and the cognition of the course of nature. Now I want to address the question of how these two ways of characterizing the distinction are related.

I think that Kant's point is that the representational resources one makes use of from each standpoint answer to the kind of purpose that one seeks to accomplish when one takes up the relevant perspective. The empirical standpoint is concerned with the aim to gain theoretical understanding of objects: an aim that includes mundane, everyday attempts to see how things hang together as well as high-level scientific theorizing. Since we are theoretically acquainted with objects only in so far as they affect our sensibility, and since our human forms of sensibility are space and in time, it follows that our theoretical understanding of objects depends on our capacity to represent them in spatiotemporal terms. This is why the standpoint of empirical cognition is also conceived by Kant as a human standpoint. We can only realize the purpose associated with occupying the empirical perspective – observing, understanding, explaining, and predicting the course of nature – by drawing on representations that have sensible content, and hence by exercising our human forms of sensibility.

From the normative standpoint, one attempts to cognize normative laws for the use of our higher faculties. It may be less obvious why pursuing this aim requires the kind of representational resources associated with the 'more than human standpoint'. We can partially answer this question by focusing first on the attempt to cognize normative laws for the exercise of the will. Here I shall refer back to the arguments I expounded in the preceding chapters. There we saw that Kant believes that we can represent agents as addressees of norm only if we can represent them as practically free and spontaneous. The idea of a free will, at least on a general level³⁷⁵, "contains nothing borrowed from experience" (CPR, A 533/B 561), and hence it is a

³⁷⁵ The general representation of practical freedom contains only the idea of the independence of the faculty of the will from determination by sensuous causes. This is a content that all rational beings are capable of representing.

'more than human' representation of pure reason that is available to all thinking beings regardless of what their forms of sensibility may be. Moreover, when engaged in practical reasoning from the normative standpoint, we consider agents not just in a more than human manner as free, but also as bound by principles that enjoy a more than human validity in that they represent justificatory standards that apply to all practically free beings. Hence, these principles (the various formulae of the categorical imperative) are essential to the practical legislation of all thinking, rational beings even if they lack human sensibility and experience. These points explain why for Kant the normative standpoint coincides with the more than human standpoint, at least with respect to normative *practical* reasoning.

Now, I have argued that Kant's normative standpoint is not *solely* a standpoint of normative practical reasoning: from this standpoint one also cognizes norms for epistemic agency. If there were a strict analogy between the case of practical and epistemic normativity, this would mean that the representation of epistemic normativity also requires that we must employ the generic 'more than human' idea of transcendental freedom for the purposes of representing cognitive agents as addressees of norms. This analogy tolerates that the concept of freedom we employ to represent the normative dimension of epistemic agency *differs specifically* from the concept of practical freedom, because the former concept does not relate to the faculty of the will. But, as I have argued previously³⁷⁶, the idea of transcendental freedom is the *general* idea of a spontaneous capacity whose exercise does not stand under determining causes in time; this general idea can be applied to different faculties to generate different kinds of freedom. When it is applied to the understanding, it designates freedom of thought or judgment: "When an appearance is given us, we are still wholly free as to how we should judge the matter. The appearance depends upon the senses, but the judgment upon the understanding" (Prol, 4: 290). Hence, the standpoint from which we cognize normative laws for the use of our higher faculties *generally* coincides with the more than human standpoint from which we represent objects as free or spontaneous, because representing an agent as an addressee of epistemic norms also requires representing her as free or spontaneous; and the representation of freedom or spontaneity is, as we saw, 'more than human'.³⁷⁷

The interpretation I am envisaging promises to provide a link between the ideas of practical and epistemic freedom (agency) via the concept of normativity. That we can, or must, regard ourselves as free agents is due to the fact that our agency has a normative dimension: and this point applies both to our practical and to our epistemic agency. Hence, interpreting Kant's standpoint distinction non-standardsly in terms of a normative-empirical contrast (rather than in terms of the dichotomy between practical and theoretical reasoning) yields the promise of providing a *unified* conception of the freedom (or spontaneity) of our higher faculties.

However, there are philosophical and interpretive problems that cast serious doubt on

However, remember (cf. chapter III, section 1) that Kant thinks that our concept of practical freedom has also some *empirical* content, in that it picks out the kinds of sensible causes that do not determine the will as sensuous causes (such as desires). Insofar as one's acquaintance with those sensible states is drawn from inner experience, it is not available to rational creatures that do not share our human forms of sensibility (i.e., time).

³⁷⁶ See my chapter III, section 1.

³⁷⁷ There is a question here about whether reasoning toward epistemic norms depends, like reasoning toward practical norms, on justificatory epistemic standards that enjoy a more than human validity. I believe that this indeed so: both the ideas of theoretical reason, and, more controversially, the unschematized categories are sources of epistemic norms that are available to all kinds of discursive thinkers. See my chapters X and XI.

whether my interpretation of Kant's standpoint distinction can live up to this promise. Let me begin here by expounding two philosophical worries that I take to be complementary to one another. First, it might be held that the practical/theoretical contrast that I am rejecting, unlike the normative/empirical contrast that I am endorsing, tracks a real distinction between two forms of reasoning. The putative contrast between empirical and normative reasoning crumbles once we acknowledge that every conclusion of a piece of empirical reasoning can be framed in normative terms: 'I ought to believe that p' (where p is some empirical fact). It would seem, then, that *all* reasoning is normative and that no informative distinction can be drawn between forms of reasoning by appeal to a normative/non-normative distinction. But my attempt to draw a distinction between normative (not merely practical) and empirical reasoning is also subject to a worry that comes from an opposite angle. It might be claimed that it is very hard to see how reasoning could be normative unless it is practical: any reasoning that is non-practical seems to concern the way things are, and although the conclusion of such reasoning may be framed in normative terms – 'I ought to believe that p' – the normative language here is reducible, or (as Richard Moran puts it) 'transparent' to, an appeal to empirical facts.³⁷⁸

This raises a dilemma for the account I am envisaging as Kant's: either the operative notion of normativity is so thin that it can be extended to the judgments that are distinctive of the empirical perspective ('I ought to believe that p') or it is so thick that it leaves nothing but practical norms for the putatively normative perspective to establish.

This philosophical worry can be supplemented with an interpretative worry. I argued that Kant's distinction between a normative and an empirical standpoint draws on the notion that we must represent agents as spontaneous to make sense of the *normative dimension* of agency. This might seem to imply that the empirical standpoint lacks a normative dimension and hence (given the envisaged link between freedom and normativity) has no use for the idea of spontaneity. But it is a crucial tenet of Kant's theoretical philosophy that the subject of empirical cognition exercises her higher faculties spontaneously and is governed by epistemic norms. Thus, there is a normative and spontaneous element in empirical cognition, and this may seem to destroy the putative contrast between a normative and an empirical standpoint.

My response to these objections crucially turns on a distinction between two senses in which a form of reasoning can be said to have a normative dimension. A piece of reasoning might be considered as *norm-governed* when it stands under normative principles or as *norm-giving* when it aims at establishing normative principles. The normative dimension of empirical cognition consists in the fact that a subject's empirical beliefs are governed by her awareness of normative laws that the agent can represent to herself in terms of the judgment that she ought to reason or judge in a certain manner. This, I want to argue, is to be contrasted with a form of reasoning and cognition that is *not only* norm-governed but (in Kant's terms) 'legislative': it aims at establishing normative principles. For the sake of elucidating this distinction between a norm-governed and a norm-giving dimension of our agency with some precision, we can try to find a common core of norm-giving, legislative reasoning *as such* that is indifferent between practical and theoretical legislation and that contrasts with the norm-governed dimension of empirical cognition. Here we should take seriously Kant's repeated insistence that the difference between

³⁷⁸ Korsgaard's contrast between the "explanatory standpoint of theoretical reason" and the "normative standpoint of practical reason" (1996a, p. 173) might be motivated by the idea that only practical reasoning is normative in any sufficiently thick sense.

the two standpoints rests, in part, on two different ways of representing or 'considering' *objects*. What Kant has in mind here is the intentional object of thought, i.e., that *about which one thinks* and that fixes the subject matter of one's reasoning. From the empirical standpoint, what one considers is a spatiotemporally located object that is governed by causal laws of nature; it is only an object of this type that can answer to the cognitive purposes associated with the empirical standpoint. One's subject matter here is a unified course of nature, a coherent spatiotemporal framework in which events happen according to explanatorily sufficient reasons. By contrast, from the non-empirical standpoint one reasons about free agents that have the capacity to act in accordance with norms. One's subject matter here is not the course of nature but what Kant calls an "order of ideas": one tries to establish a conception of how things ought to happen that is independent of how things actually happen in nature. So Kant's striking claim that "ought has no meaning when one has only the course of nature in view" (CPR, A 547/B 575) should be taken literally, as relating to *what one has in view*, namely, the intentional object of thought (or reasoning) from the empirical standpoint: a causally determined object that one *must not* represent in terms of the more than human idea of freedom or spontaneity, and correspondingly a descriptive subject-matter about which we cannot intelligibly raise normative questions.

The idea to distinguish between reasoning from the normative and the empirical standpoint by saying that the former kind of reasoning has a distinctively normative *subject matter* helps to avoid the first horn of the above dilemma. We can indeed say that empirical reasoning is reasoning about what we should believe about the empirical world; but empirical reasoning is still to be considered as non-normative in the deep sense that we cannot intelligibly raise normative questions about its subject matter. Another way of putting this point is by saying that in empirical reasoning the normative fact about what the *subject* of thought ought to believe can be identified with, or is transparent to, a non-normative fact about the *object(s)* of thought. By contrast, in a case of non-empirical, normative reasoning, the normative fact represented by the judgment that an agent ought to act in a certain way cannot be identified with a non-normative, descriptive fact about (say) the way the agent does or did or will act. Hence, there is an informative sense in which only conclusions of reasoning from the normative standpoint can be deemed *irreducibly normative*: they, unlike conclusions of empirical reasoning, are not 'transparent to' descriptive conclusions about empirical, natural (non-normative) facts.

But now the second horn of the above dilemma becomes relevant. If what is distinctive of irreducibly normative reasoning is that it does not concern natural facts, or that it does not coincide with reasoning about the course of nature, then (it might be held) there is *one and only one* kind of reasoning that satisfies this condition: normative *practical* reasoning. Any non-practical normative reasoning, concerning, for instance, how a subject of empirical cognition should judge is identifiable with reasoning about the way the empirical world is: it depends on empirical observation and is answerable to natural facts. This, it might be said, is why a genuinely normative standpoint can only be a standpoint of practical reasoning.

This objection brings us to the heart of the matter, not just of Kant's standpoint distinction but also of his account of epistemic normativity. The interpretation of his standpoint distinction that I am envisaging depends crucially on an idea that, I think, is not just of interpretative but also of wide-ranging philosophical significance: namely, the idea that there is a normative question, 'how ought one to judge the natural world?', that cannot be answered by observing, predicting and explaining how things are in nature, but that must be answered *prior to*

and independently of any actual attempt to observe, predict or explain the course of nature. If so, it makes sense to identify a form of irreducibly normative reasoning that aims at cognizing *norms for* empirical cognition without thereby aiming at empirical cognition (and hence without collapsing into empirical reasoning); and then we can contrast an irreducibly normative but not merely practical standpoint with a standpoint of empirical cognition.

Now, I think that Kant is indeed committed to there being norms of empirical cognition that must be established prior to and independently of empirical cognition. Consider, for instance, the norm that "prescribes that we ought to study nature *as if* systematic and purposive unity, combined with the greatest possible manifoldness, were everywhere to be met with, *in infinitum*" (CPR, A 701/B 729). Kant holds that this necessary principle *cannot* be established from the empirical standpoint "through observation of the accidental constitution of nature" (CPR, A 651/B 679). Notice that this principle employs one of the ideas of theoretical reason (namely, the idea of systematic unity) that Kant mentions in the passage from the Groundwork where he delineates the non-empirical standpoint of pure spontaneity. This shows that Kant does not hold that theoretical reason is concerned only with what is as opposed to what ought to be the case.³⁷⁹ Now, this conception of epistemic normativity, as deriving from the purely spontaneous exercise of theoretical reason, raises many questions, some of which I shall address at a later point (in chapter XI); and there is a question here about whether *the understanding* also has a non-empirical, normative use (a question to which, as I shall argue in chapter X, Kant gives a very special affirmative answer that is shaped by his idealism). In anticipation of my arguments in subsequent chapters, I want to suggest that Kant's conception of epistemic normativity is crucially informed by the point that our quest for empirical cognition stands in need of criteria for empirical truth and of principles that organize the quest for such truth; criteria and principles that cannot, on pain of circularity, in turn be gained on the basis of empirical cognition.

For our present purposes, the important upshot of the preceding discussion is that we can draw the distinction between the two standpoints in terms of what I shall call their object conception: from the normative standpoint, the intentional object of reasoning (what we reason about) are agents that we characterize as norm-governed, free, spontaneous, and sensibly affected (but not determined), that is, in terms of concepts all of which are 'more than human'. The addition 'sensibly affected' is important here: when establishing practical norms, we picture the addressee, i.e., the practical subject of volition, as affected by sensible desires; and when establishing epistemic norms, we picture the addressee, i.e., the epistemic subject of cognition, as affected with sensible data. (This is why we represent norms as *oughts*.) From the empirical standpoint, the object of reasoning is a natural object that we characterize, in 'merely human' terms, as spatiotemporally located and as causally determined by natural states.

I want to suggest that we can also draw the distinction between the two standpoints in terms of what I shall call their subject conception, namely, by appeal to the kind of *self-awareness* that we are afforded with when occupying the two standpoints. From the normative standpoint, the subject of reasoning is aware of herself as *purely* spontaneous, that is, as producing representations without being in any sense constrained by the deliverances of her passive, sensible nature.³⁸⁰ From the empirical standpoint, the subject of reasoning is aware of

³⁷⁹ See Bojanowski 2006, p. 169, p. 216, for the idea that theoretical reason has no normative aspect. Allison 2004, p. 331, p. 446 and Korsgaard 1996a, p. 204, p. 218 are sensitive to the normativity of theoretical reason; but their interpretation of Kant's standpoint distinction seems, on the whole, oblivious to this point .

³⁸⁰ This is why Kant emphasizes "...those [representations] that we produce simply from ourselves, and in which we

herself as both spontaneous and receptive, i.e., as being dependent, for the exercise of her spontaneity, on the deliverances of her passive, sensible nature.³⁸¹ These two different kinds of self-awareness are, for Kant, connected to a difference in the (awareness of) the validity and authority of the representations that we produce, respectively, from the two standpoints: the judgments that we arrive on the basis of purely spontaneous reasoning that borrows nothing from sensibility or experience articulate *laws* that enjoy a universal, a priori validity and that express a strict necessity; whereas the judgments that we arrive at from the empirical standpoint articulate mere *rules* that, in virtue of their dependency on experience, have only comparative validity.³⁸²

I have argued that Kant contrasts a normative standpoint that involves reasoning toward practical and epistemic norms with a standpoint of empirical cognition. But I want to conclude my argument by conceding that there is an important disanalogy between 'legislative' reasoning toward practical norms and toward epistemic norms: in the epistemic case, the 'execution' of the relevant norms requires a switch of perspective from the normative to the empirical standpoint, for it consists in the formation of judgments about the empirical world. In the practical case, this execution takes place from the normative standpoint itself, when one exercises one's will to form maxims in the light of previously established normative insights.³⁸³ But this difference should not detract from the fact that the conception we have of ourselves qua subjects of reasoning from the normative standpoint is, in a deep sense, the same regardless of whether we reason toward

show our own activity..." – representations deriving from "pure activity...which reaches consciousness immediately and not through affecting the senses" (GMS, 4: 452). Compare CPR, A 546-9/B 574-77.

³⁸¹ We are aware of "our liability to be affected by objects" (CPR, B 42-3) and of "representations given us from without, and in [respect to] which we are passive" (GMS, 4: 451-2).

³⁸² These rules are empirical concepts. While these concepts "are formed according to law" (i.e., according to the epistemic laws articulated from the normative standpoint), "they are not themselves legislative, but the rules founded on them are empirical and, consequently, contingent" (KU, 5: 174). This does not cohere with all of Kant's immensely varied commitments. The schematic picture, laws: normative standpoint of pure spontaneity versus rules: empirical standpoint of sensible affection faces two counterexamples. First, in the epistemic case, Kant thinks that there are some *necessary laws* that can be established from the empirical perspective despite a (limited) dependency on empirical data: namely, the physical principles he examines in the *Metaphysical Foundations*. Here Kant should at least admit that these principles have not as strong a claim to necessity and universality as those epistemic laws that are wholly independent of experience. Second, in the practical case, Kant's distinction between laws and rules is that between categorical and hypothetical imperatives (GMS, 4: 389), and here the matter is quite complicated. In order to generate mere practical rules, one must indeed occupy the empirical standpoint of causal reasoning (to find, on inductive grounds, means toward happiness); but the material provided by empirical cognition takes the form of a practical rule only if practical reason supplies its *sui generis* concept of conditional value and represents the relevant course of action as conditionally good, which requires the normative standpoint. See my chapter IV, section 2 for discussion of this issue.

³⁸³ Notice, however, that practical legislation is not *essentially* tied to the executive function of the will: practical legislation is neither intrinsically forward-looking (we may aim at making true moral judgments for the purposes of passing blame or praise on *past* actions) nor intrinsically first-personal (we may aim at making true practical judgments for the purposes of giving advice to others). I shall return to this important point in the appendix. Notice also that if one moves from practical legislation to practical execution, one loses the sense of the purity of one's spontaneity, for the faculty of Willkür that is responsible for volition is always sensibly affected. Only practical *reason* is capable of a *purely* spontaneous exercise. For the distinction between pure practical reason and sensibly affected (but not determined) Willkür, see MS, 6: 226 and MS, 6: 213: "The human Willkür is in fact affected by [senuous] impulses or stimuli, but is not determined by them; and it is, therefore, not pure in itself...." In exercising our capacity for legislative practical reasoning, we are (aware of ourselves as) members of the intelligible world; in exercising our sensibly affected faculty of Willkür, we are (aware of ourselves as) members of both the intelligible and the sensible world. See chapter III, section 3 for further discussion.

practical or epistemic normative laws, i.e., regardless of whether we exercise our higher, purely spontaneous practical or epistemic faculties. It is a conception of ourselves as *legislators* and hence as *autonomous* beings, "in respect of the faculties of the soul generally, insofar as they are regarded as higher faculties, i.e., as faculties containing an autonomy" (KU, 5: 196).

5: The relation between the two standpoints: distinctness and interdependence

Now that we have a sense of what Kant's standpoint distinction amounts to, I want to characterize the relation between the two perspectives and their characteristic enquiries.

The first thing to notice here is that the two standpoints seem to be *distinct* in the following two senses. First, as we saw, occupying the two stances means representing and 'considering' objects in distinctive and contrasting ways, namely, as addressees of norms who are free from causal determination (from the normative standpoint) and as natural, spatiotemporally located objects that are determined by causal laws (from the empirical standpoint). Second, there is a corresponding difference in the kinds of reasoning one engages in from the two standpoints. While there are many different species of empirical and of normative reasoning, they all share a common general core: no kind of empirical reasoning is constrained, in its attempt to arrive at a conclusion about its respective object, by the extent to which this object stands under normative principles; and no species of normative reasoning can realize its attempt to arrive at a conclusion about its respective object by deferring to descriptive, predictive or explanatory facts about its object. The following examples may help to make this abstract point more concrete. Consider the proposition, 'John kills his brother at time t1'. One can approach this proposition in a predictive spirit, and ask whether John will indeed kill his brother at t1, or at least how likely is it that he is going to do it. But since John is a rational agent, one can also take an entirely difference stance toward the same proposition and ask whether he should kill his brother at t1. Likewise, regarding the statement, 'Caesar crossed the Rubicon in 49 BC', one can disregard what Caesar should have done and try to disclose what the historical facts of the matter are, or what explains those facts; on the other hand, one can disregard the question of what psychological and social-historical conditions serve to explain the fact that Caesar crossed the Rubicon in 49 BC, and instead address the question of whether it was a good thing that he acted as he did.

The fact that normative and empirical reasoning are distinct does not preclude the further suggestion that they are also *interdependent* – an idea that might, in its most ambitious version, amount to the claim that there is a 'division of labor' between these two forms of reasoning.³⁸⁴ This would be the case if it were true, first, that normative reasoning depends on the deliverances of empirical reasoning and, secondly, that empirical reasoning depends on the deliverances of normative reasoning. The first condition is satisfied at least in the case of normative practical reasoning: what ought to be done depends on what is the case in the empirical world (for example, whether or not I ought to help a friend depends on whether he is in need of help) and what lies in people's physical power.³⁸⁵ The second condition is also satisfied if one can defend

³⁸⁴ GMS, 4: 388: "All trades, arts, and handiworks have gained by division of labor, namely, when, instead of one man doing everything, each confines himself to a certain kind of work distinct from others in the treatment it requires, so as to be able to perform it with greater facility and in the greatest perfection. Where the different kinds of work are not distinguished and divided...manufactures remain still in the greatest barbarism. (...) I only ask here whether the nature of science does not require that we should always carefully separate the empirical from the rational part..." 'The rational part' is what I consider the normative part.

³⁸⁵ Even though the primary object of practical laws is the volitional activity of people, to be in a morally required

Kant's claim – that I sketched only briefly – that empirical cognition presupposes normative principles that are not themselves established on the basis of empirical cognition.

Is there a conflict between the idea that normative and empirical reasoning are distinct and yet interdependent? We can partly defend the coherence of this idea by showing that practical normative reasoning seems distinct, *at each step*, from empirical reasoning. There are two relevant possibilities here: the first occurs when we take for granted empirical knowledge that we already possess. For instance, I tell my friend that he should have picked up the phone to tell me that he will be late, and I take for granted that he was in fact late, that he knows how to use the phone, etc. Here, I rely on previous deliverances of empirical reasoning, but I do not therefore *engage in* empirical reasoning. My reasoning is not *about* whether my friend has the capacity to use phones. There is, therefore, a sense in which normative reasoning brackets empirical concerns: while empirical facts of various kinds can be relevant to normative reasoning, normative reasoning is never concerned with the attempt to resolve empirical questions. Now – and this is the second possibility – normative reasoning may face a question about what the empirical facts are where no previous deliverances of empirical reasoning can be drawn upon. For instance, consider the question of whether a man who witnessed someone drown is morally responsible for not jumping in the water to help. The answer to this question will depend partly on whether the man can swim. If I do not know whether he had the capacity to swim, normative reasoning requires that I first resolve this factual question. In such a case, we temporarily switch from one viewpoint to the other. While I try to ascertain whether it was physically possible for the man to save the drowning person, I occupy the empirical perspective; once I have settled this question, I switch back to the normative perspective and consider what, in the light of my newly acquired empirical knowledge, the man should have done.³⁸⁶

But there *is* a problem with the idea that the two standpoints are both distinct and interdependent that arises because of Kant's specific understanding of what a standpoint is. As we saw, for Kant the distinction between the two standpoints is tantamount to the distinction between a 'merely human' manner of representation and a 'more than human' manner of representation. Now one might raise the following objection: (1) There could only be a 'more than human' standpoint that is distinct from a 'merely human' standpoint *if the former does not make any use* of representations that are available solely to beings with human forms of receptivity. (2) The more than human standpoint is a standpoint of normative reasoning, and (at least practical) normative reasoning inevitably involves a reference to empirical facts that are represented according to specific forms of sensibility. (3) Hence, the 'more than human', normative standpoint cannot be truly distinct from the 'merely human', empirical standpoint.

(2) is correct. (1) might be based on Kant's claim that "it is only from the human

state of willing is to intend certain physical effects in the natural world: from a first-person point of view, willing to help a friend means doing all in one's physical power to help her. See my chapter VII, section 3.

³⁸⁶ One might hold that if one finds out that the man could not have saved the other person from drowning, this settles the question of whether he ought to have saved the person from drowning. This would conflict with the claim that no empirical cognition of what the world is like can settle a question of normative reasoning. But the reasoning from, 'The man could not have saved her' to 'It is not the case that he ought to have saved her' is not itself a piece of empirical reasoning; that one cannot stand under a normative prescription to do something that is beyond one's physical power is itself a norm that cannot be established, interpreted, or applied on the basis of a piece of empirical cognition. Moreover, as we saw (in chapter VII, section 3), establishing that the man could not have saved the other person does not at all settle the question of whether he complied with or violated the relevant moral norm at the level of volition.

standpoint that we can speak of space, of extended things, etc" (CPR, A 26/B 42). Notice, however, that this claim cannot be meant literally: even from the more than human standpoint, we can "speak of" spatiotemporal properties in a certain way – namely, *as* transcendently ideal, as *not* being characteristic of the constitution of objects when 'they are considered in themselves through reason.' Kant's remark should be read as a hyperbolic way of saying that spatiotemporal "predicate[s] can only be ascribed to things in so far as they appear to us, that is, to objects of sensibility" (CPR, A 26-7/B 42-3). If our aim is to represent the constitution that objects have independently of the conditions of our sensibility, and hence to represent them in a manner available to all rational beings, we cannot characterize *this constitution* in spatiotemporal terms: we cannot consider beings as they are in themselves according to spatiotemporal characteristics. But certainly Kant does not think that we somehow forget the fact that those beings are *also* objects of human sensibility and hence have an ontological character that conforms to the conditions of human sensibility.³⁸⁷ That is, even from the more than human standpoint one *can* acknowledge the phenomenal properties of objects, as long as one does not represent those predicates as pertaining to the 'more than human' character of the object that one *has in view* from this standpoint (a character whose existence enables the existence of a more than human perspective of representation). We must keep apart here the following acts: (a), acknowledging that objects also have an empirical character; and (b), 'considering them' *according to* their empirical character. The difference between (a) and (b) parallels the difference between merely drawing, within normative reasoning, on one's empirical knowledge and trying to arrive at such knowledge through empirical reasoning. It is (b), not (a), that is incompatible with occupying the more than human perspective, because in (b) but not in (a) one's thinking about the object *must answer to and is constrained by* the conditions imposed by human forms of sensibility.

The distinction between (a) and (b) is also relevant to the natural question of how one can consider an action *in time* (and hence according to a merely human form of sensibility) from the normative, more than human standpoint; for instance, when one passes blame on a past action. The point is, again, that from the normative standpoint, we can acknowledge the temporal position of actions (as established by empirical observation). But we cannot (from the normative, more than human standpoint) *consider actions according to* their temporal properties, that is, we cannot try to assign temporal positions to acts and relate them (in terms of concepts such as duration, precedence, or succession) to other events within a coherent temporal framework: if we did, we would depart from the more than human, normative standpoint, for we would consider actions according to a merely human form of sensibility (time) and as mere parts of the deterministic course of nature about which we cannot intelligibly raise normative questions.

The conclusion we should draw from these considerations is that the notion of a more than human perspective from which we consider beings normatively, according to 'ideas' (of reason), involves a certain idealization. Kant's point cannot be that a thinking being who occupies this perspective is completely oblivious to her empirical knowledge of the world of sense and hence to the empirical representations of which she is capable by virtue of exercising her special (for instance, human) faculties of sensibility. Rather, his point here is twofold. First, there is a representational 'core' that is absolutely essential to the normative purposes associated

³⁸⁷ Kant *must* allow that we can make a reference to the spatiotemporal world even when we think of objects as things in themselves, because he wants to make room for the notion that our atemporal, non-spatial agency has effects that unfold in space and in time: "reason...presupposes that it can have causality in regard to all these actions, since otherwise no empirical effects could be expected from its ideas" (CPR, A 548/B 576).

with the more than human standpoint, which consists in representing beings as free, spontaneous, rational and norm-governed, that is completely independent of the deliverances of sensibility, and that is (hence) available to all thinking beings. Second, while the kind of reasoning one engages in from this standpoint must draw on the deliverances of empirical cognition of the sensible world, such reasoning does not aim at empirical cognition of the sensible world – it does not, in Kant's terms, have the course of nature 'in view' as an object of enquiry and hence is not (unlike empirical cognition) constrained by what can be represented and observed in sensible experience. This lack of constraint comes to light in the conviction that rational beings can act on the basis of nothing but respect for the moral law even if sensible experience cannot provide any instance of such motivation; or in the verdict that people ought to act in a certain manner even if experience shows that they constantly fail to act in this manner; or in the idea that rational beings have a distinctive kind of value (dignity) that cannot be displayed in sensible experience.

In my view, this shows that we can reject premise (1) of the above objection and hold that the integrity of a more than human perspective on beings and their actions can tolerate the fact that one draws (within strict limits) on representations supplied by a merely human perspective.

But there is a second way in which one might press the worry that Kant is not entitled to the claim that the two standpoints are both distinct and interdependent. Kant thinks that the distinctness of the more than human from the merely human perspective comes to light in the fact that the laws that are legislated from the former perspective, at least the practical laws, are valid for all rational creatures and not merely for human beings (GMS, 4: 389). But normative practical judgments can only issue concrete prescriptions by making a reference to determinate empirical circumstances, and this requires the use of spatiotemporal representational resources. One might worry that the intrusion of 'merely human' forms of representation into the 'more than human standpoint' negates the distinctness of the two points of view by limiting the scope of the validity of the normative principles that one legislates from the more than human standpoint.

To address this worry, we must get clear about what Kant means when he claims that the verdicts of practical reason are not only valid for human beings. He says (my emphasis):

As my concern here is with moral philosophy, I limit the question suggested to this: Whether it is not of the utmost necessity to construct a pure thing which is only empirical and which belongs to anthropology? for that such a philosophy must be possible is evident from the common idea of duty and of the moral laws. **Everyone must admit that if a law is to have moral force, i.e., to be the basis of an obligation, it must carry with it absolute necessity; that, for example, the precept, 'Thou shalt not lie,' is not valid for men alone, as if other rational beings had no need to observe it; and so with all the other moral laws properly so called; that, therefore, the basis of obligation must not be sought in the nature of man, or in the circumstances in the world in which he is placed, but a priori simply in the conception of pure reason.** (GMS, 4: 389)

This passage shows that Kant equates the universal, more than human validity of moral laws with their *strict necessity*, that is, with the notion that such laws state what has to be done.³⁸⁸ Necessary practical laws must not have their *foundation* or *basis* in "the nature of man, or in the circumstances in the world in which he is placed". Instead they must have their basis in pure reason. What Kant is concerned with here is the standard of justification to which specific practical prescriptions owe their authority. This standard must derive from an idea which has its

³⁸⁸ Compare GMS, 4: 408: "...unless we deny that the notion of morality has any truth or reference to any possible object, we must admit that its law must be valid, not merely for men but for all rational creatures generally, not merely under certain contingent conditions or with exceptions but with absolute necessity..."

basis in pure reason alone: i.e., the categorical imperative must inform any specific judgment which purports to prescribe with strict necessity.

If this is what is required for the 'more than human validity' of a practical law, the fact that reasoning toward such laws depends on premises concerning how things are in the 'merely human' spatiotemporal world does not vitiate the more than human validity of practical laws. To see this, let us take a piece of practical reasoning that (arguably) issues in the legislation of a practical law. There are people in Africa who are threatened by death from starvation; this is a fact about the human world of sense. Given this fact, there is (arguably) a piece of practical reasoning that leads to the verdict that I am morally required to give money to those people rather than spending it on a new Ipod. Although the correctness of this reasoning does depend on an empirical premise (if there was no starvation in the world, it might be permissible for me to spend my money on luxuries), and although the resulting prescription essentially concerns what is to be done in the spatiotemporal world, the *foundation or basis* for this reasoning and its verdict is independent of anything empirical. The verdict prescribes with strict rational necessity because it is legitimized by justificatory standards (such as the standard that represents humanity, but not the pleasure I take in Ipods, as an unconditional good or end) that are rooted in nothing but pure practical reason. Kant might say that other finite rational beings would be subject to an *exactly analogous* concrete prescription, that is, one that makes *the same moral point* by drawing on whatever empirical concepts these other beings would use, in accordance with their forms of sensibility, to represent analogous empirical circumstances.

It must be admitted that Kant tends to insist on the idea that practical principles are valid not just for human beings precisely with respect to principles that are *so general* that they can indeed be arrived at and formulated without appealing to any spatiotemporal, empirical facts: this includes both the formulations of the categorical imperative itself³⁸⁹ and principles that are somewhat more specific than the general formulations of the moral law but, arguably, still general enough to prescribe without employing empirical concepts – principles such as, 'Thou shalt not lie'.³⁹⁰ But his demand for the more than human validity of practical principles upon closer examination always and without exception reduces to the demand that those principles prescribe with strict necessity and hence cannot be founded on empirical, psychological aspects of human nature; this demand, as we saw, *is* compatible with allowing that the prescriptions refer to specific circumstances in the natural world that are represented in spatiotemporal terms.³⁹¹ Thus, we should conclude that when Kant states his claim that practical laws are valid not just for human beings, he has in mind no more than the demand that such laws are founded on justificatory standards that are rooted in pure practical reason (a faculty we share with other rational beings) and that do not borrow anything from the contingent sensible conditions that

³⁸⁹ GMS, 4: 430-1: "This principle, that humanity and generally every rational nature is an end in itself (which is the supreme limiting condition of every man's freedom of action), is not borrowed from experience, firstly, because it is universal, applying as it does to all rational beings whatever, and experience is not capable of determining anything about them."

³⁹⁰ Here I am assuming that the concept of a lie is not drawn from experience, which may be controversial.

³⁹¹ The psychological aspects of human nature *do* provide the basis for rules of prudence or happiness, and *hence* Kant denies that these rules can be regarded as genuine practical laws (GMS, 4: 389-90). This shows, again (cf. footnote 382 above) that the idea that the normative standpoint is a 'more than human standpoint' is only a general, programmatic point that does not cohere with all of Kant's varied and complex commitments: mere practical rules are, despite their lack of more than human validity, normative.

affect our empirical, sensible nature.³⁹² Again, the idea that the normative standpoint qualifies as 'more than human' involves a certain idealization but nonetheless seems justified in virtue of the fact that this standpoint has a certain essential core (here: of justificatory standards) that is completely independent of anything that is merely human.

I want to address a third way of pressing the worry that the interdependence of the two standpoints conflicts with their distinctness. One might object that from the allegedly non-empirical standpoint, which is supposed to abstract from the fact that our phenomenal acts are governed by causal laws of nature, one must (in figuring out what one ought to do) take into account facts about what lies in one's physical power and what is empirically possible.³⁹³

There are various ways of fleshing out this worry. First, it might concern the fact that in order to arrive at a specific conclusion about what one ought to do, one must have a causal understanding of the empirical world. In response to this point, I would reiterate that the mere fact that a thinker *draws on* causal facts established by empirical reasoning does not impugn the idea that she is *not engaged in* empirical reasoning and hence does not occupy the empirical perspective. If one is focused on the question of what causal means are required for realizing a certain end, one is not engaged in normative reasoning. (Kant clearly thinks that causal means-end reasoning qualifies as a theoretical or empirical form of enquiry.³⁹⁴) Of course, one may realize, in the course of normative practical reasoning about what to do, that one must figure out causal means toward certain ends, and on this basis switch from the normative to the empirical perspective (and back again). But this is consistent with holding that one does not occupy the empirical and the normative standpoint simultaneously.

But the worry might also be understood as follows: if figuring out what we ought to do requires an awareness of what it is physically possible for us to do and hence of the fact that our agency stands under causal laws of nature, then there is no genuine sense in which we are abstracting from the idea that we are causally determined; but wasn't the idea that we abstract from the fact that we are so determined somehow integral to the notion of a standpoint from which we are conscious of our membership 'in the intelligible world'? My response here draws on an important point that I expounded in some detail elsewhere (chapter VII, section 3): Kant's idea of practical freedom relates only to our will, not to our body. His claim that from the normative standpoint we abstract from the fact that we are causally determined is meant to pertain only to our conviction that we can exercise our noumenal powers of practical reasoning and choice in 'inner acts', independently of causal determination: we wholly abstract from the fact that our phenomenal powers of choice are causally determined by psychological laws and conditions. Although we expect that our noumenal choices may have the intended effects in the physical world of appearance, this is not something we can know or safely rely on, precisely because we must acknowledge that our phenomenal bodily powers are constrained by empirical conditions beyond our control. This explains Kant's insistence that the *only* practical act that always lies in every rational agent's power is the formation of a pure maxim.³⁹⁵

³⁹² Whether an analogous point holds for epistemic normative laws remains to be seen (cf. my chapter X).

³⁹³ See Nelkin 2000 and Watkins 2004, p. 322 for variations of this worry.

³⁹⁴ This is clear from his remark that providing for 'technical-practical rules' requires only theoretical cognition, but not a specifically practical branch of philosophy. See KU, 5: 172 and EEKU, 20: 199-200. See my chapter IV, section 2 for further discussion of this point and of the complications that it raises.

³⁹⁵ See KpV, 5: 36-7. See my chapter VII, section 3 for a discussion of how this point affects Kant's conception of moral responsibility.

6: The compatibility of the two standpoints

Undoubtedly, the most important question concerning the relation between the two standpoints is how they could be regarded as compatible, given the fact that from the normative standpoint we represent human agents as free whereas from the empirical standpoint we represent them as causally determined. In this section I want to discuss the attempts of three contemporary philosophers to reconcile the empirical/scientific and the normative conception that we have of ourselves by appeal to a distinction between two standpoints. I then want to point out how Kant's picture differs from each of these three attempts. This will allow us to see how Kant takes his standpoint distinction to *partly* legitimize our self-conception as free, norm-governed agents.

We can begin by looking at Peter Strawson's account in "Skepticism and Naturalism". He claims that we should make ascriptions of freedom and responsibility only when we occupy what he calls the 'engaged' standpoint, from which one can consider an action "as an appropriate object of a reaction of moral disgust"; and he argues that the very same action can "rightly [be] seen as merely the natural outcome of a complex collocation of factors, an appropriate object of scientific, psychological and sociological analysis and study."³⁹⁶ Strawson wants to admit that the latter way of viewing the action requires one to deny that the agent is free or responsible.³⁹⁷ Yet, he claims that from the 'engaged' standpoint, we are rationally justified in believing that agents are free and responsible.³⁹⁸ He thinks that the claims, 'we are free and responsible' and 'we are not free and responsible' do not conflict because their content includes a reference to the different standpoints from which these claims are made, which renders each of these claims true *relative to their respective standpoint*.³⁹⁹

Now, the notion of a claim's (or it's content) being *relative to a standpoint* is rather elusive, and this obscurity may well indicate that Strawson's account is ultimately unintelligible.⁴⁰⁰ Rather than pressing this point, I want to expose a further difficulty that I take to be fatal for Strawson's account. Note that he is not making the compatibilist claim that someone

³⁹⁶ Strawson 1987, pp. 41-42.

³⁹⁷ Strawson does not merely make the trivial claim that naturalistic explanations do not employ concepts of freedom and responsibility. His point is that those explanations *exclude* the legitimacy of such ascriptions: they imply that the relevant concepts are not instantiated (see his 1987, pp. 49-50).

³⁹⁸ In "Skepticism and Naturalism" (1987), Strawson is not merely saying that it is psychologically unavoidable that we apply concepts of freedom and responsibility from the 'engaged' standpoint. The appeal to the psychological unavoidability of conceiving of ourselves as responsible agents is meant to play an argumentative role in "Freedom and Resentment" (2003). The problem here is that with respect to every action, it is possible for *some* deliberator to occupy the 'naturalistic standpoint' and to expose another deliberator's 'engaged' ascription of responsibility as illegitimate. And if this kind of verdict were correct, there would always be a standpoint that some deliberator can occupy from which a given ascription of responsibility is exposed as illegitimate, even if no human being has the psychological capacity to occupy this standpoint all the time with respect to every ascription of responsibility. In his 2003, Strawson has an argument for the claim that such ascriptions are actually not threatened by determinism: he claims that incompatibilism about responsibility derives from an illegitimate generalization of ordinary exemptions and excuses (see my chapter VI, section 2). But this argument is not repeated in 1987. Another argument which plays no role in 1987 is Strawson's third compatibilist argument from 2003, the argument that we would have practical reasons to view ourselves as responsible agents even if theoretical reason told against the possibility of responsibility. Strawson's appeal to two different standpoints in 1987, as I understand it, is meant to go beyond each of the three flawed compatibilist arguments from 2003.

³⁹⁹ Strawson 1987, p. 38: the claim that we are free and responsible must be understood as "relative to the standpoint which we normally occupy as social beings".

⁴⁰⁰ See Stroud's 1999, pp. 183-7 trenchant criticism of Strawson's appeal to standpoint-relativity.

who thinks that there is a conflict between the belief that some action is blameworthy and the belief that the action is the result of deterministic processes is operating with a mistaken conception of responsibility. If that were his position, he should invoke the proper (compatibilist) conception of accountability; but he rather puts the entire argumentative weight on the suggestion that each of the above beliefs must be relativized to a standpoint. Now suppose that R believes (1) that the principle of alternative possibilities (which states that an agent is responsible for her actions only if she could have acted otherwise) is legitimate, and (2) that the relevant sense of possibility requires freedom from causal determination. Suppose further that R believes (3) that determinism is true so that some action *x* causally determined. It would seem that by accepting (1)-(3), R cannot accept (4) that agent T is responsible for action *x*.

Strawson seems to suggest that R can accept (4) if only she relativizes (4) to the 'engaged standpoint'. But (1)-(3) imply that (4) is false, and making what Strawson calls 'the relativizing move' does nothing to cancel this implication. Someone who accepts (1)-(3) is committed to thinking that no one has alternative possibilities and thus that an ascription of responsibility (like (4)) must be false. The rational pressure to reject (4) that is created by (1)-(3) cannot be dissolved by the relativizing move, because that move does not bear on what (1)-(3) imply. Since the tension between (3) and (4) arises in virtue of the acceptance of (1) or (2), it could be resolved by the rejection of either (1) or (2). But if the tension *is* thus resolved, we need not relativize (4) to some standpoint. So the 'relativizing move' turns out to be either superfluous or impotent.

Strawson might respond that my rejection of the legitimizing force of the 'relativizing move' depends on my claim that this move is needed in the first place because of the acceptance of (1) or (2) or both. He might hold that this claim need not be accepted, because (1) and (2) exhibit substantive theoretical commitments that cannot simply be presupposed. But this response is not convincing. Strawson must concede that the 'relativizing move' is necessary *in the first place* because of some tension between (3) and (4). If there were no such tension, the 'relativizing move' would be superfluous, and one could make ascriptions of responsibility even from the 'detached naturalistic standpoint' from which one gives deterministic explanations. If there *is* such a tension, then one must ask *why* (3) and (4) seem to be in conflict, and the natural answer is that this is because of the acceptance of something like (1) and (2).

This shows that the appeal to the idea that the claim that we are free and responsible is made from a distinctive standpoint cannot *suffice* to legitimize this claim: the putative legitimizing force of the appeal to the notion of a standpoint dissolves once we examine how the need to legitimize this claim arises in the first place.

A different strategy for showing that an appeal to standpoints can reconcile apparently conflicting aspects of our self-conception is suggested by Hilary Bok. Unlike Strawson, Bok does not suggest that we can render compatible the claims, 'I am free to do *x*' (advanced from the practical standpoint) and 'I am not free to do *x*' (advanced from the theoretical standpoint) by relativizing them to their respective standpoints. On Bok's account, these claims do not conflict because they employ *different concepts of freedom or possibility* that are tailored to the distinctive purposes associated with the standpoints. From the theoretical standpoint, and thus for the purposes of explanation and prediction, we use the notion of 'possibility *tout court*', which signifies what we can do given all relevant information about our state, circumstances, antecedent causes, and so forth; not abstracting from any of this information is required by the

purpose of theoretical reasoning.⁴⁰¹ By contrast, the concept of freedom and possibility we use when we adopt the practical standpoint is what Bok calls 'possibility in a general sense', which allows that one abstracts from certain types of information to specify what is possible. More precisely, Bok claims that from the practical standpoint it is legitimate to abstract from information concerning the causal conditions that regulate how someone will *choose*. Abstracting from this information gives us, she suggests, the conditional notion of possibility ('it is possible for R to do x' = 'R would do x if R chose to do x'):

If we define our possibilities thus [by disregarding facts about the causal conditions of choices], then we can say that all of those actions that we would have performed had we chosen to perform them are possibilities for us, just as we can say that it is possible for a given amount of water to vaporize if we disregard the fact that it is sitting in my freezer.⁴⁰²

I must admit that I find it very hard to see what the point of Bok's analogy is supposed to be here. She is right to emphasize that we can affirm that the water in my freezer can vaporize only if we abstract from the fact that it is sitting in my freezer. But surely, we can affirm that if R had chosen to do y, she would have done y without 'disregarding' the fact that R is causally determined not to choose to do y. This points to a crucial difficulty in Bok's account: her insistence that we can legitimately use the general, conditional notion of possibility only from the *practical* standpoint seems puzzling because it seems that establishing whether something is possible in this conditional sense is a task for *theoretical* reasoning. This task requires that one does *not* abstract from theoretically relevant information: if I want to know whether R can, in the conditional sense, get out of bed, I must not abstract from any information concerning R's physical abilities (e.g., I must consider whether R is paralyzed). Hence, what is possible in the conditional sense is also possible 'tout court'. The two concepts are not in conflict. Likewise, for settling the question of whether R would get out of bed if he chose to do so, it is *theoretically*, not merely *practically* irrelevant whether or not R can choose to do so.⁴⁰³

Now, Bok suggests that from the practical standpoint we have reason to attribute to ourselves not only (conditional) freedom of action, but also freedom of choice:

...if it is in principle impossible for an agent to know, before making a choice, what she will choose to do, then no conceivable correction of her beliefs could allow her to narrow the set of actions that she regards as alternatives beyond those that she would perform if she chose to perform them. (...) The conception of our possibilities that we use to define the alternatives among which we can choose must be the compatibilists' general conception of possibility, rather than the narrower possibility *tout court*. (...) When we deliberate, we take the question what we will...choose to do to be open – not because we believe the outcome of our choice to be physically undetermined, but because we regard the question what we will in fact choose to do as one whose answer depends on us, and which we have yet to answer. (1998, p. 108; pp. 114-5)

⁴⁰¹ Bok 1998, p. 94.

⁴⁰² Bok 1998, p. 94.

⁴⁰³ Bok suggests that "the claim that an agent would have performed some action had she chosen to do so has no particular importance from a theoretical point of view. What would have happened had some condition obtained does not figure in an explanation of what happened when that condition did not obtain" (1998, p. 104). But the conditional notion of 'can' tracks the physical capacities of people, and one can surely have a purely theoretical interest in what the physical abilities of people are – whether, for instance, someone would get out of bed if she chose to do so. Moreover, it seems clear that facts about what a person could have done if she had chosen to do must figure in a theoretical explanation of what she did, because what it is physically possible for people to do (and what *they believe* to be possible for them to do) is in the background of an explanation of what they did.

This argument (whose logic and adequacy I have discussed at some length in chapter VI) suggests that Bok thinks that the concept of *freedom of choice* that we should use from the practical standpoint is that of epistemic possibility – we must take the question of what we will choose as epistemically open because 'it is in principle impossible for an agent to know what she will choose before making a choice'. But, again, the idea that our knowledge of our future choices *is* thus limited is a *theoretically respectable* notion: we can – indeed, must – declare that our choices are epistemically open to us from the theoretical standpoint itself.

Bok's account promises to avoid the clash between our theoretical and our practical conception of our freedom (of what is possible for us to do) by emphasizing that the practical and the theoretical standpoint involve different concepts of freedom or possibility. But, as it turns out, Bok thinks that the only concepts of possibility we have reason to use from the practical standpoints are compatibilist concepts (conditional freedom of action, epistemic freedom of choice). The theoretical standpoint does not in the least call into question (but rather *establishes*) that we are free in those compatibilist senses. If it is conceded that the only concepts of freedom we are concerned with when reasoning practically are compatibilist concepts, then it follows immediately that there is no conflict between our theoretical and our practical self-conception: compatibilist concepts of freedom are *as such* theoretically (naturalistically) acceptable, and there is no sense of 'free' according to which the claim, 'we are free' would be denied by theoretical and affirmed by practical reasoning. Indeed, as we saw, theoretical reasoning itself must be called upon to find out which of our actions we are free, in the conditional sense, to perform. But then the distinction between two standpoints is completely superfluous for the purposes of legitimizing our practical claim to our freedom. The entire argumentative weight in Bok's position is carried by the claim that the incompatibilist concept of possibility, which signals freedom from causal determination, is irrelevant to our practical self-conception.

The most famous contemporary attempt to legitimize our self-conception as free agents by invoking a distinction between standpoints is made by Christine Korsgaard. However, one can discern several lines of thought in her work that do not obviously cohere.

One such line of thought seems to center on the suggestion that the difference between the two standpoints is a difference in the *kinds of attitudes* one takes toward the idea that one is free from causal determination. From the theoretical standpoint, one ought to take the cognitive attitude of *believing* that one is *not* free in this sense. But from the practical standpoint, adopting a cognitive attitude is a complete non-issue: "the point is not that you must *believe* that you are free, but that must choose *as if* you were free"⁴⁰⁴. It is worth noticing that the logic of *this* appeal to standpoints presupposes that one accepts an incompatibilist conception of freedom but does not invoke any metaphysical resources that go beyond a naturalistic, deterministic ontology. If one accepted either a compatibilist conception of freedom or a metaphysical framework according to which deterministic laws do not exhaustively constrain our capacities, there would be no need to insist that we must adopt a non-cognitive attitude toward the idea that we are free from causal determination when we make our choices. The problem is that Korsgaard does not tell us what, precisely, the practical attitude is that we are supposed to adopt toward the proposition that we are free from causal determination. None of the familiar practical attitudes (such as desire, intention or choice) is a meaningful candidate here: Korsgaard cannot intend the

⁴⁰⁴ Korsgaard 1996a, p. 162. On the other hand also says things such as, "we must believe these things in order to obey the categorical imperative" (1996a, p. 163; pp. 175-6); see my chapter V, section 3.

absurd suggestion that we choose *to be free* from causal determination, but rather implies that *while making our choices* we must adopt some *further* practical attitude toward the content, 'my choice is free from causal determination'. But the character of that further practical remains elusive. Linguistically, the idea that we must choose 'as if' we were free might suggest that while choosing we must *imagine* ('for the sake of the choice') that we are free. But this would completely trivialize the debate about free will. No one denies that we are entitled to imagine that we are free from causal determination.

A different strand in Korsgaard's work suggests a picture that comes rather close to Bok's account. Korsgaard argues that the fact that we are causally determined could *only* interfere with our freedom of will or agency if it could give us knowledge about what we will do that prevents us from making choices and from acting on the basis of choice.⁴⁰⁵ Since we cannot know how we will act in a way that preempts our deliberative choice, it follows that causal determinism does not threaten our practical freedom. The idea that we can choose 'as if' we were free from causal determination may then be taken as a quirky way of expressing the point that it is *irrelevant to* our freedom of choice whether or not we are causally determined to choose and act as we do.⁴⁰⁶ If this is Korsgaard's picture, then the impact of the appeal to standpoints is as limited as it is on Bok's picture. The real argumentative weight is carried, first, by the compatibilist point that the metaphysical fact that we are causally determined is irrelevant to our freedom as long as we cannot cognize how we will act in ways that would interfere with our agency; and secondly, by the *theoretical* point that such cognition is impossible. There is nothing we can say about our freedom from the practical standpoint that we cannot also say from the theoretical standpoint. The fact that the two standpoints do not conflict is due entirely to the fact that the operative concept of freedom is a compatibilist, theoretically respectable concept.

The most radical strand in Korsgaard's writings suggests that the 'external' perspective we take on ourselves when we engage in theoretical observation misses something to which we have genuine, (quasi-) perceptual or intuitive access when we adopt an 'internal' or 'first-personal' perspective toward various phenomena:

Value, like freedom, is only directly accessible from within the standpoint of reflective consciousness. (...) From [an] external, third-person perspective, all we can say is that when we are in the first-person perspective we find ourselves to be valuable, rather than simply that we are valuable. (...) Trying to actually see the value of humanity from the third-person perspective is like trying to see colors someone sees by cracking open his skull. From outside, all we can say is why he sees them. (...) If you think colors are unreal, go and look at a painting by Bellini or Olitski, and you will change your mind. (1996b, pp. 124-5)

There is no doubt that Korsgaard wants to make the exact same point about freedom that she makes about colors. But the analogy does not carry over. One may try to resist the inference from the fact that scientific explanations of color perception do not need to mention color to the idea that therefore color is not real. But even if *that* inference can be resisted, the case of freedom is

⁴⁰⁵ Cf. Korsgaard 1996b, p. 95: "The worry seems to be that if we were sure we were determined or knew how we were determined then either we could not act or we would not act..."; this is the "practical problem that knowledge could somehow take away our freedom". See again, my chapter VII, section 1.

⁴⁰⁶ This interpretation is suggested by Korsgaard's claim (1996a, p. 163; here she is addressing someone who knows that her decisions are causally determined by a brain implant): "you will believe that your decision is a sham, but it makes no difference... in order to do anything, you must simply ignore the fact that you are programmed, and decide what to do – just as if you were free."

very different. Either the operative concept of freedom is a compatibilist concept, in which case, as we saw, a theoretical perspective on things *establishes* that we are free in this sense. If, on the other hand, the operative concept of freedom designates the absence of causal determination, then it is not only true (as in the color case) that scientific explanations *do not mention* our freedom from determination by natural causes – rather, the sciences, by adducing causal factors that explain our choices and actions, give us explicit reason to believe that our choices and actions are *not* free in this sense. It seems, then, that this third version of Korsgaard’s appeal to standpoints *reinforces* the clash between the theoretical or the practical perspective.

The discussion of Strawson’s, Bok’s and Korsgaard’s accounts is helpful for the purposes of situating Kant’s own account. Kant does not endorse Strawson’s thought that the reason why the claims, (1) 'Humans are free' and (2) 'Humans are causally determined' do not contradict one another has to do with the fact that these claims are relativized to a standpoint (i.e., that the concept of a standpoint figures in their content). He never (unlike, on one reading, Korsgaard) denies that the attitude we take toward (1) from the normative standpoint is one of full-fledged belief. He never suggests that the concept that we use from the normative standpoint to represent our freedom is a compatibilist concept (this is Bok’s and, on the second reading, Korsgaard’s account). Finally, he rejects the idea (Korsgaard’s third suggestion) that we have quasi-intuitive access to our freedom from a 'first-personal perspective'.

What, then, *is* Kant’s solution? We must first remember here⁴⁰⁷ that Kant thinks that one can reconcile the claims that objects are both *spatial and non-spatial* if one recognizes that these claims represent objects under two different subject concepts:

If we add the limitation of a judgment to the concept of the subject, then the judgment is unconditionally valid. The proposition ‘All things are next to one another in space’ is valid only under the limitation that these things be taken as objects of our sensible intuition. If here I add the condition to the concept and say: ‘All things, as outer intuitions, are next to one another in space’, then this rule is valid universally and without limitation. Our expositions accordingly teach the reality (i.e., objective validity) of space in regard to everything that can come before us externally as an object, but at the same time the ideality of space in regard to things when they are considered in themselves through reason, i.e., without taking into account the constitution of our sensibility. (CPR, A 27-8/B 43-4)

We avoid a contradiction between, 'Objects are spatial' and 'Objects are non-spatial' by thinking of the objects, respectively, as appearances or as things in themselves.⁴⁰⁸ It is clear that Kant seeks to extend this point to the putative conflict between viewing one and the same being as free and as causally determined:

Nevertheless, even though we should never be able to comprehend how freedom is possible, we must at least remove this apparent contradiction in a convincing manner. For if the thought of freedom contradicts either itself or nature, which is equally necessary, it must in competition with physical necessity be entirely given up. It would, however, be impossible to escape this contradiction if the thinking subject, which seems to itself free, conceived itself in the same sense or in the very same relation when it calls itself free as when in respect of the same action it assumes itself to be subject to the law of nature. **Hence it is an indispensable problem of speculative philosophy** to show that its illusion respecting the contradiction rests on this, that we think of man in a different sense and relation when we call him free and when we regard him as subject to the laws of nature as being part and parcel

⁴⁰⁷ See my chapters I and II for elaborate discussions of this point.

⁴⁰⁸ Notice that the idea that we represent objects under two different subject concepts does not at all imply that we represent them as two *numerically different objects*. See my chapter I for my defense of this idea.

of nature. It must therefore show that both [senses] very well co-exist.... (...) **This duty, however, belongs only to speculative philosophy.** (...) There is not the smallest contradiction in saying that a thing in appearance (belonging to the world of sense) is subject to certain laws, of which the very same as a thing or being in itself is independent, and that he must conceive and think of himself in this twofold way, rests as to the first on the consciousness of himself as an object affected through the senses, and as to the second on the consciousness of himself as an intelligence, i.e., as independent on sensible impressions in the employment of his reason (in other words as belonging to the world of understanding). (GMS, 4: 456-7; my emphasis.)

Now where determination according to laws of nature ceases, there all explanation ceases also, and nothing remains but defense, i.e., the removal of the objections of those who pretend to have seen deeper into the nature of things, and thereupon boldly declare freedom impossible. We can only point out to them that the supposed contradiction that they have discovered in it arises only from this, that in order to be able to apply the law of nature to human actions, they must necessarily consider man as an appearance: then when we demand of them that they should also think of him *qua* intelligence as a thing in itself, they still persist in considering him in this respect also as an appearance. In this view it would no doubt be a contradiction to suppose the causality of the same subject (that is, his will) to be withdrawn from all the natural laws of the sensible world. But this contradiction disappears, if they would only bethink themselves and admit, as is reasonable, that behind the appearances there must also lie at their root (although hidden) the things in themselves, and that we cannot expect the laws of these to be the same as those that govern their appearances. (GMS, 4: 459)

Making room for the possibility of endorsing both the claim that we are free and the claim that we are causally determined is, according to Kant, entirely up to 'speculative', i.e., theoretical philosophy. Theoretical philosophy must demonstrate that there is a *metaphysical space* for the existence of freedom alongside that of determinism, and that there is logical space for the belief that we are free alongside the belief that we are causally determined: those who dogmatically deny our freedom pretend to see further 'into the nature of things' as the metaphysical framework of transcendental idealism permits.⁴⁰⁹ This basis is provided by the defense of transcendental idealism in the Critique of Pure Reason, where Kant takes himself to establish: first, that the principle of causal determination is both a condition on the possibility of human experience and on the objects of human experience; secondly, that the principles of possible experience are not absolute metaphysical principles of being that pertain to the unknowable constitution that objects exhibit independently of our cognitive faculties. Under these assumptions, there is metaphysical space for assuming that objects have an ontological character that differs from the character they have *qua* objects of human experience. This does not, by itself, provide us with positive reasons for thinking that we are free.⁴¹⁰ But this is not what is *currently* at issue in my discussion. The present concern is with the question of how the claims that we are free and that we are unfree (causally determined) could both be affirmed without contradiction. As the above passages show, Kant's appeal to standpoints is not meant to replace attempts to make ontological space for the assumption that we are free from causal determination. Rather, his appeal to standpoints *presupposes* that there is such ontological space and hence that there is no contradiction in affirming that we are noumenally free and phenomenally determined by natural causes.

But it is important to see that it does not follow from the fact that for Kant the appeal to standpoints does not itself remove this contradiction that therefore his appeal to standpoints plays *no* role in legitimizing our self-conception as free agents. Affirming that we are free would be illegitimate if it contradicted the true claim that we are causally determined; this clash is avoided

⁴⁰⁹ See my chapter II, section 3 for an elaboration and defense of this interpretation.

⁴¹⁰ See my chapter II, section 3, and chapter XIII.

if one accepts the metaphysical framework provided by transcendental idealism. But notice that representing objects as free from causal determination can be illegitimate *even within* this framework. It is helpful here to distinguish between the metaphysical possibility that we are free and the theoretical credibility and fruitfulness of the idea that we are free. Transcendental idealism provides a framework within which freedom is metaphysically possible, but it rules out the idea that this possibility could ever be shown to be actual on the basis of standards that we rely on in empirical theorizing about objects: freedom cannot be an object of intuition or perception. Moreover, for precisely this reason, the concept is fruitless for the purposes of theorizing about what objects are like, or for the purposes of observing, predicting, or explaining things, not just in the sense that it cannot further our knowledge of objects or events but in the more radical sense that it violates the conditions (such as temporality) for having objective experience of objects and events. Hence, it is very important for Kant to show that we apply the concept of freedom to objects only from a standpoint from which we engage in a legitimate form of enquiry that is not aimed at empirical cognition and whose standards for applying concepts are not fixed by empirical evidence. It is not an objection to a form of reasoning that seeks to establish *what ought to happen* that it operates with a concept that cannot be used for explanatory or observational purposes (and that even violates the conditions of objective experience).

Here the contrast between Kant's picture and the picture endorsed by Bok (and, on one interpretation, Korsgaard) is illuminating. On the latter picture, the concepts of freedom we use for normative (practical) reasoning are not only metaphysically but also theoretically credible: that we have conditional freedom of action (with respect to certain kinds of action) and that we have epistemic freedom of choice is confirmed by our theoretical understanding of ourselves as parts of the natural order. This is what makes it so hard to see what work the appeal to distinctive standpoints is supposed to be doing on this picture. We can legitimately apply all the relevant concepts of freedom or possibility from *any* standpoint. By contrast, for Kant we *cannot* apply the concept of freedom from causal determination to objects if we take the perspective of empirical cognition toward objects and consider them in accordance with our forms of sensibility. This is why for him the idea that we use this concept from a non-empirical, more than human standpoint is an indispensable aspect of his legitimization of that concept.

People are often mystified when they hear that all it takes to combine the idea that we are free from causal determination with the idea that we are causally determined is the suggestion that these ideas are advanced separately from distinctive standpoints.⁴¹¹ I agree that merely appealing to different standpoints cannot have the magical impact of removing a contradiction between a pair of beliefs. On my interpretation of Kant's standpoint distinction, this distinction does play a genuine role in legitimizing our representation of human agents as free (in that it legitimizes the use of an empirically empty idea by appeal to the fact that this idea is used from a standpoint that does not have to answer to the standards of empirical cognition), but Kant does not impose a burden on this distinction that it could not possibly bear.

7: Summary: the significance of Kant's standpoint distinction

In this chapter I attempted to get clear about what Kant's standpoint distinction amounts to and what work it is doing in his defense of the legitimacy of the claim that we are free. I take the intuitive core of Kant's standpoint distinction to be obvious and appealing. It is in fact the case

⁴¹¹ See Nelkin 2000 for a very perspicuous articulation of why this looks like a mystery.

that we can take two distinctive perspectives on human agents: we can view their behavior from an observational, explanatory, or predictive perspective, as the causal upshot of natural conditions and processes; or we can view their behavior from a normative perspective as answerable to rational principles. That these are two distinctive ways of considering ourselves and others is a pre-philosophical 'datum'. Kant's standpoint distinction should be regarded as a (very involved) philosophical interpretation of this datum. He gives an in-depth characterization of (first) the representational resources that we must bring to bear on our conception of objects from the two standpoints and (second) of the kind of awareness we have of ourselves as subjects who occupy the empirical and the normative stance.

If my interpretation of Kant's standpoint distinction is correct, then what makes his way of interpreting the intuitive contrast between the normative and the empirical perspective especially interesting is the fact that he does not think that the normative perspective relates one-sidedly to the practical aspect of our agency; rather, Kant seems to accept that we can also consider our *cognitive efforts* in two ways, as natural events and as answerable to rational, normative considerations. When we consider an agent's cognitive efforts as subject to normative considerations, we must represent the agent as having freedom of thought or judgment from causal determination. This suggests a picture on which there is a tight link between freedom (or spontaneity) and normativity. More precisely, the link is twofold and relates to two dimensions of our agency: first, to the legislative, *norm-giving* aspect of our agency, which is to be conceived in terms of the idea of a *pure* kind of freedom or spontaneity (whose exercise is independent of conditions of sensible affection); and secondly, to the executive, norm-governed aspect of our agency, which is to be interpreted in terms of the idea of a kind of freedom or spontaneity whose exercise is affected, but not determined, by the influence of sensible conditions. Normativity qua normative legislation and qua norm-governedness implies, and is implied by, the idea of freedom or spontaneity. On this interpretation of Kant's standpoint distinction, there is a way to connect up his talk of spontaneity in the practical sphere with his talk of spontaneity in the theoretical sphere. Rather than treating practical and theoretical spontaneity as two isolated and separable notions, my interpretation yields the promise of relating them to one another via the concept of normativity – or rather, via the two dimensions of normativity as normative legislation and as norm-governedness. We are practically and theoretically free insofar as we are the legislators (sources) as well as the executors of normative prescriptions.

My interpretation may seem to invite an important objection. I have interpreted Kant's standpoint distinction (partly) in terms of the distinction between two forms of reasoning, and one might worry that the latter distinction is not exhaustive: for instance, it cannot accommodate the important case of pure mathematical reasoning, which is neither normative nor empirical. In fact, however, considering objects 'mathematically' does not fit into Kant's standpoint distinction quite independently of my interpretation of that distinction. Kant's standpoint distinction *must* be understood as a distinction between considering things as 'intelligible' objects and as empirical objects. In pure mathematical reasoning, one considers objects as spatial, and thus not as 'intelligible'; but one does not consider them as object of empirical intuition (and as being governed by causal laws) either. Now there are two possible replies to this worry. First, it seems that the mathematical standpoint is closely tied to the empirical standpoint, for three reasons: both the empirical and the mathematical perspective are 'merely human' in that they consider objects according to merely human forms of sensibility; both perspectives have a decidedly

descriptive subject matter, which means that we cannot intelligibly raise normative questions about the intentional objects of mathematical and empirical reasoning⁴¹²; and the standpoint of pure mathematical reasoning is itself answerable to, or depends for its legitimacy on, the standpoint of empirical cognition that applies mathematical concepts to the natural world of sense.⁴¹³ The second response grants that the perspective of pure mathematics is in a deep sense independent of the empirical perspective but denies that this is problematic for Kant: it would be problematic for Kant only if his ambition were to comprise all kinds of reasoning activities and, correspondingly, all forms of representing objects within one standpoint distinction. If he has no such ambition, he can simply concede that (in the terms of my interpretation) the normative and the empirical standpoint do not exhaust the space of possible forms of reasoning and of representing objects. Moreover, he can even give a principled reason for leaving out of account a potential standpoint of pure mathematical reasoning. His official standpoint distinction is motivated, to a large extent, by the goal to show that we can adopt two distinctive perspectives *on human agents*: either as addressee of normative laws or as being governed by empirical, causal laws. The possibility of pure mathematical reasoning does not provide a counterexample to the claim that Kant's official standpoint distinction is exhaustive with regards to the perspectives we can adopt with respect to human agents.⁴¹⁴

8: Appendix: Kant's standpoint distinction versus Kantian standpoint distinctions

I now want to relate Kant's standpoint distinction, as I have interpreted it, to the standpoint distinctions that have been drawn by some prominent Kantians in recent years. It will turn out that the differences between Kant's own standpoint distinction and the Kantian conceptualizations of the distinction reflect substantive philosophical disagreements.

The first picture I want to compare and contrast with Kant's is a familiar one according to which the essential difference between two standpoints concerns the difference between practical and theoretical reasoning. I have argued that this is too narrow as an interpretation of Kant's picture, because what Kant contrasts with the empirical standpoint is a normative

⁴¹² This is why Kant's contrast between normative and empirical reasoning is at the same time a contrast between normative and mathematical reasoning (CPR, A 547/B 575): "It is just as absurd to ask what ought to happen in the natural world as to ask what properties a circle ought to have. All that we are justified in asking is: what happens in nature? What are the properties of the circle?"

⁴¹³ See CPR, A 157/B 196 for the claim that the synthetic a priori judgments of pure mathematics depends for their legitimacy on their mediate relation to possible sensible experience.

⁴¹⁴ There is a further complication here: we may view ourselves not just as governed by either normative or causal laws but also as subject to *teleological* standards – namely, when we view ourselves as natural organisms. I take the teleological perspective to be an enlargement of the originally mechanistically oriented empirical standpoint. Notably, when we consider ourselves according to the kinds of teleological standards and concepts that are proper to biological explanation as Kant conceives of it, we can think of ourselves under *norms* of functioning, but I do not think that this signals an unacceptable intrusion of normativity into the empirical perspective. The kind of normativity that derives from the application of teleological standards is fundamentally different from the prescriptive kind of normativity that I have appealed to in delineating the distinction between the normative standpoint (as a perspective from which we prescribe how rational beings ought to act) and the empirical standpoint. We do not have to consider ourselves as free or rational in order to consider ourselves according to teleological standards and norms: after all, teleological norms also apply to blades of grass, and the necessary laws of practical reason and freedom that Kant is concerned with are categorically different from the model of means-end rationality that allows for the purposive explanation of the actions of human beings and higher animals by appeal to belief-desire pairs.

perspective as such which includes both reasoning toward norms of volition and reasoning toward epistemic norms. But even if we leave out of account this issue and focus exclusively on the case of normative practical reasoning, a crucial difference between Kant's picture and (say) Korsgaard's picture remains. This difference concerns the question, 'what is practical reasoning?' Some philosophers literally define practical reasoning by appeal to its putative conclusion: the formation of intention or choice.⁴¹⁵ Others hold that the perspective of practical reasoning is exclusively a forward-looking perspective from which an agent attempt to arrive at a decision about what she should do.⁴¹⁶ Both of these proposals are in conflict with Kant's official picture because, as we saw, in Kant the faculty of practical reason that produces moral laws is conceptually distinguished from Willkür as the faculty that is responsible for our choices. The exercise of practical reason is not necessarily tied to the exercise of Willkür.

I think that Kant's distinction between two separate practical faculties is well-motivated. Consider first the suggestion that practical reasoning concludes in the formation of an intention. Suppose that I am aware that by telling a 'harmless' lie to my partner, I can avoid frustrating discussions; and suppose I go through a process of reasoning in which I move, first, from the acceptance of (say) the proposition that lying for the sake of hedonic benefit amounts to treating others as mere means to the proposition that lying to my partner would be morally forbidden, and, secondly, from the latter proposition to the judgment, 'I ought not to lie to my partner'. Now there is a *separate* question of whether I end up choosing not to lie on the basis of my recognition that lying is morally wrong. Suppose I give in to the temptation to tell the lie. Here my decision to lie, and my act of lying, are wrong; but there seems to be no tension between voicing such a criticism and conceding that at least my practical *reasoning* has been correct. The relevant criticism concerns my willpower, but not the exercise of those rational capacities that provide me with a correct judgment about what I ought to do. Moreover, once a practical judgment such as 'I should not lie to my partner' has been formed, it seems that *no further piece of reasoning* could effect a transition to a state of mind that could be regarded as the determination not to lie. Reasoning consists in the reasoned transition from a proposition or set of propositions to another proposition; and once I have accepted the proposition that I should not lie to my partner, there is no transition to be made to the acceptance of a further proposition.⁴¹⁷ What is needed now is not a further act of reasoning that aims at the truth but an exercise of willpower.⁴¹⁸ This provides a rationale for dividing the will into two sub-faculties: one

⁴¹⁵ See Broome 2009. For similar proposals, see Velleman 2000 and Bratman 1987. The fact that all these philosophers identify practical reasoning by appeal to intention as its conclusion-state should not mask the fact that they all disagree about what, exactly, an intention is; notoriously, Velleman holds (whereas Bratman denies) that an intention is to be treated as the deliberator's *prediction* about what she will do.

⁴¹⁶ See Bok 1998, p. 62. This is also Korsgaard's position; contrary to Broome and others, she denies (in her 2009) that we have any need to posit intentions as the conclusion of practical reasoning.

⁴¹⁷ Broome 2009 seems to hold that it is possible to literally reason toward the determination of one's will, for he thinks that the content of the attitudes one adopts in practical reasoning includes, besides propositional content, "a mark of some sort": for instance, the mark of a desire-attitude is a word such as 'nice', and presumably, the 'mark' of the attitudes one adopts through practical reasoning ensures that one's will is determined to act in the way specified by the propositional content of the attitude. Broome's talk about 'markers' is elusive, and Broome does not specify what the 'mark' of the characteristic conclusion of practical reasoning (for him, intention) is supposed to be (he says that it is *not* the imperative mood, and he talks of 'inaudible markers', but this merely adds to the obscurity).

⁴¹⁸ Raz 2010 offers some closely related criticisms of the idea that practical reasoning can be distinguished by appeal to the fact that it terminates in an intention or choice. However, Raz concludes that it follows that there is

responsible for the legislation and one for the execution of practical laws.

Now, there is a further question here about whether the exercise of the faculty of practical reason is solely concerned with answering the question, 'what should I do?' If so, this would imply that one reasons practically only if one raises normative questions about *one's own future actions*. Kant would, it seems to me, reject this 'narrow-range' conception of practical reasoning, because he defines practical reason as a law-giving faculty that is exercised whenever a moral law is established; and this can be in cases where one considers past actions as well as in contexts where one considers the (past or future) actions of other people. There are two intuitions that speak in favor of such a 'wide-range' conception of practical reasoning. First, one's reasoning in contexts where one gives others advice about what to do or in contexts where one holds oneself or others responsible for what they have done is governed by the same standards of correctness (for Kant, the formulae of the categorical imperative) as reasoning about what oneself should do in the future. Second, in these latter contexts one represents the agent one reasons about as having practical freedom: a kind of reasoning that makes use of this representation surely has a solid claim to qualifying as practical.⁴¹⁹

One strategy for defending the narrow-range view is to argue that cases where we reason about what we or others should have done or cases about what others should do are *covertly* forward-looking and first-personal. What comes to mind here is Allan Gibbard's claim that, as Barry Stroud puts it, "all questions of what we or anyone else ought to do are planning questions; arriving at a conclusion about what ought to be done is really a matter of adopting a plan".⁴²⁰ The idea here is that all judgments about what ought to be done reflect an agent's attempt to work out what Gibbard calls a "contingency plan"; having such a plan is to be reflectively prepared to choose and act under all kinds of circumstances.⁴²¹ Now, this proposal seems untrue to the 'phenomenology' of (say) giving advice. When I reason to the conclusion that my friend ought not to cheat on his partner, it does not at all *seem to me* as if I am engaged in an episode of planning, i.e., as if I am, at bottom, concerned with *my* decision in potential circumstances of temptation: my concern is with *him* and his *actual* plight. This appeal to the 'phenomenology' of giving advice to others is not decisive. But if we can come up with a case where a person reaches what looks like a meaningful verdict of practical reason that *cannot possibly* be conceived as a covert contingency plan, this would decisively tell against Gibbard's conception of practical reasoning. Consider a man who knows that he is going to die within a

no distinction at all to be drawn between practical and theoretical reasoning, and I think that we can continue to hold that practical reasoning is distinguished from other forms of reasoning by its subject matter – the articulation of norms for how we should exercise the *will*.

⁴¹⁹ Korsgaard, a defender of the narrow-range conception of reasoning, seems to agree that if one's reasoning makes use of the idea of freedom, it qualifies as practical; but she thinks that it is only in first-personal forward-looking reasoning about what one should do that one has use for the idea of freedom: "Freedom is the capacity to do otherwise, not the capacity to have done otherwise" (1996b, p. 96). By contrast, Kant thinks that "all men attribute to themselves freedom of will. Hence come all judgments upon actions as being such as ought to have been done although they have not been done" (GMS, 4: 455). Their disagreement about the contexts in which we can consider ourselves free rests on a further, deep disagreement about what threat determinism poses to our freedom, and about how we should respond to this threat. I have expounded this fundamental difference between Korsgaard (who thinks that it is enough to show that we cannot know in advance of making our choices what we will do, and that we have epistemic freedom of choice) and Kant (who thinks that the threat determinism poses to our agency is metaphysical and that this threat can only be answered in metaphysical terms) in chapter VI.

⁴²⁰ See Stroud's 'Introduction' to Gibbard 2011, p. 5

⁴²¹ See Gibbard 2011, p. 170; p. 177.

matter of minutes. In a final act of reflection, he realizes that he ought to have paid more respect to his wife. Assuming that this conclusion is correct and arrived at on a sound deliberative route, it seems that the man has reached *an insight* – his practical reason has made him understand something about how people ought to treat one another. And yet the man cannot be said to be engaged in an episode of future contingency planning. He knows that he has no future.⁴²²

This brings us to what I think is the fundamental root of the disagreement (between Kant and Gibbard/Korsgaard) about what counts as practical reasoning. Gibbard seems drawn to his identification of practical reasoning with planning because he cannot see how normative attitudes would be of any relevance unless they were adopted for the purpose of helping us figure out what to do; and this worry rests on his diagnosis that unless they are connected to our planning efforts, normative beliefs cannot be understood as being about facts. Likewise, Korsgaard worries extensively about the 'realist' idea⁴²³ that moral reasoning aims at giving us knowledge of facts; this idea, she fears, would render unintelligible the practical relevance of moral principles – a relevance they could only have when they arise in the context of first-personal, forward looking reasoning about what to do.⁴²⁴ Now it should be clear that Kant does not think that normative practical reasoning tries to apprehend a set of facts that are on a par with natural facts about trees and atoms. But he does think that there *is* something to be understood and learned in practical reasoning: practical reasoning, *at least in its pure form*, attempts to depict a rational, normative order of how things ought to happen (or ought to have happened).⁴²⁵ When such an attempt succeeds, we have not acquired a piece of knowledge about the constitution of natural objects in the world of appearances, but we have learned something about what it takes to display proper respect for the dignity of rational beings. When it is asked: 'what is the relevance of this understanding, independently of attempts to determine our own future conduct?', we need only look at the case of the dying man who eventually understands that and

⁴²² Since one philosopher's modus ponens is another one's modus tollens, might Gibbard just respond that *since* the man is not developing a contingency plan, he cannot be reasoning practically? This would be to beg the question against someone who denies that practical reasoning must be a matter of contingency planning. When we compare the dying man's reasoning toward the conclusion that he ought not have done a certain thing with the reasoning about the same subject matter by someone who *has* a future, we can see that their reasoning is governed by the exact same standards. This seems to make a good case for the idea that the dying man's reflection qualifies as practical reasoning; it is Gibbard who must convince the defender of the wide-range conception of practical reasoning that this is a mistake.

⁴²³ Korsgaard 1996b, p. 35. I think that this worry relates only to what she calls 'substantive realism'.

⁴²⁴ See Korsgaard 1996b, p. 37; p. 17. Gibbard does not deny that normative attitudes are genuine beliefs and concern matters of fact (see Gibbard 2011, p. 179). What he denies is that this classification allows us to understand the point and function of normative attitudes: "We might have thought that we needed ought beliefs to figure out what to do. On an approach that identifies them as beliefs and leaves it at that, however, they don't seem needed" (ibid., p. 178).

⁴²⁵ With respect to the articulation of *norms of prudence*, I think that Kant would not doubt that the main relevance of one's verdict that acting in a certain manner was imprudent derives from the effect it is taken to have on our future planning – i.e., the regulation of our own future happiness. The truth of this verdict depends to a large extent on standards of theoretical reasoning (see my chapter V, section 2) whose relevance for us depends on the interest we take in our own happiness – and this interest, unlike the interest we have in what it takes to properly respect the value of humanity, does seem to be exclusively forward-looking. However, it should be noticed that giving others advice concerning what they, non-morally, ought to do does *not* have to be related to one's own planning efforts – it can be a moral duty to prevent others from doing things that (as one has theoretical reason to believe) would make them unhappy.

why it was wrong to disrespect his wife.⁴²⁶ For the man, this insight counts – he has understood something that is *worth knowing*. To picture practical reasoning as essentially aimed at planning one's future conduct loses sight of the notion that understanding one's status and duties as a moral agent can be an end worth attaining in and by itself, independently of the extent to which this relates to our status as planning creatures. For Kant, practical reasoning has a significance that does not derive exclusively from its action-guiding role; this explains why Kant holds that moral reasoning aims at *cognition* of moral laws.⁴²⁷ That such cognition is possible requires not that we cognize values like we cognize heavenly bodies, but that there are intersubjective standards by appeal to which our moral reasoning can be assessed as correct or incorrect, and that moral principles can be regarded as objective in the sense that their validity does not depend on the contingent psychological states that may or may not affect human beings. For Kant, these conditions are satisfied even if our moral reasoning does not relate to our own, future actions.

We can conclude from the preceding paragraphs that for Kant, normative practical reasoning cannot simply be identified with forward-looking deliberation about what to do. It is aimed at the articulation of principles concerning how rational beings ought to determine, or ought to have determined, their will: this reasoning is called 'practical' because it concerns the working of a practical faculty and hence has a practical subject matter. Of course it is *also very important* for Kant to show that pure practical reason can (by itself) determine our will and to move us to act. But it simply does not follow that we must say that when we reason toward moral principles in a context where the determination of our own will is a non-issue, we are covertly constructing a contingency plan.⁴²⁸ This concludes my discussion of the difference between Kant's standpoint distinction and the more common distinction between a standpoint of practical and theoretical reasoning. For Kant, the standpoint from which we engage in practical reasoning is *not solely* a perspective from which we address the question, 'what should I do?' The different conceptualizations of the distinction between two standpoints in Kant and (e.g.) Korsgaard reflect a deep-seated disagreement about the nature of practical reasoning, the relevance of ascriptions of freedom, and the connection between freedom and responsibility.

⁴²⁶ We could also appeal to cases where we reason toward practical norms in the context of holding others responsible: here, too, a judgment about what ought have happened has a significance that is not reducible to the significance it may have for how people will act in the future. Those who identify practical reasoning with its forward-looking, planning function must attempt to read a planning, future-directed attitude into all contexts where we hold ourselves and others responsible for what we or they have done. But (as Wallace 1994 shows) this ignores the backward-looking dimension of judgments of responsibility, which explains our commitment to the idea that the agent we are holding responsible had the capacity – the freedom – to act differently. If we made those judgments solely in a forward-looking spirit (e.g., with the aim of making sure that things will go as they should in the future), then all we ought to care about is that an agent whom we praise or blame can be causally manipulated, through our praise and blame, to act as we desire in the future. This point is well-brought out by those consequentialists who reject the idea that agents could have acted differently in any relevant sense, but who defend the logic of holding others responsible by appeal to the idea that (e.g.) blaming a person might increase the likelihood of acting differently in the future (cf. Smart 2003). If one thinks that this causal interpretation of the point of our practice of holding people responsible is misguided, one should also be skeptical about the attempt to interpret this practice exclusively in terms of a forward-looking attitude.

⁴²⁷ See, for instance, GMS, 4: 410; 452; and CPR, A 823/B 851.

⁴²⁸ It should also be noticed that in the famous passage where Kant argues that we recognize our practical freedom on the basis of our awareness of moral laws, his example does *not* involve a case where a person seeks to form a choice or intention, but one where a person considers what she ought to do in a *purely hypothetical* scenario where she *would* face the choice between telling a lie or being condemned to death.

I now want to turn to a second standpoint distinction that one often finds associated with Kant, namely, that between a first-personal and a third-personal perspective.⁴²⁹ For Korsgaard, this contrast is simply identical to that between theoretical and practical reasoning. A more interesting version of the first-personal/third-personal distinction is offered by Richard Moran. Although Moran, like Korsgaard, contrasts the 'deliberative' (= first-person) with a 'purely theoretical' (= third-person) standpoint, he thinks that from the deliberative standpoint one is concerned with the formation of both 'practical' attitudes (intention or decision) *and* 'theoretical' attitudes (belief).⁴³⁰ His distinction is intended to track the difference between cases in which a subject's attitudes are sensitive to her conception of whether there is good reason to hold them and cases where her attitudes exist or persist regardless of her conception of whether these attitudes are justified: for instance, when a weak-willed agent predicts that she is not going to abide by her decision to stop gambling; or when a patient in a Freudian setting learns that she holds an unconscious belief that she was abused by her father. It is a virtue of Moran's distinction that he does not suppress the normative dimension of belief formation: in forming beliefs, we take ourselves to be responsive to what we think are good reasons for believing one thing rather than another about the world, and we exercise our rational capacities. I take this to be very congenial to Kant's account, which (as we saw and are going to see) also portrays the formation of empirical beliefs as essentially norm-governed and agential. However, it is important to see that Kant's standpoint distinction cannot be between a first – and a third-personal perspective.

First, it is essential to Moran's conception of the 'first-person deliberative perspective', as opposed to the 'third-personal theoretical' perspective, that holding or abandoning some attitude (mental state) is sensitive to one's judgment about whether there are good reasons for having that attitude. But for Kant, to consider the attitude of *another* person – and hence to consider that attitude third-personally – is not at all to adopt a theoretical or empirical perspective. From Kant's non-empirical, normative standpoint, one may, for instance, take up the question of what maxims some other person should adopt or what she should have done.

Moreover, it is important to see that Moran's distinction – and, it seems to me, the first/third person distinction quite generally – only ever applies *relatively to a given attitude*; absolutely speaking one never ceases to occupy Moran's first-person deliberative perspective. In the therapeutic setting where R attempts to realize whether she believes that p, R's question about whether she believes that p is not 'transparent to' the question of whether p is the case. Thus her belief that p is not governed by the characteristic norms of belief-formation, where such governance is, Moran holds, the distinctive mark of the first-person deliberative perspective.⁴³¹ But it cannot be inferred that R is not occupying this perspective *at all*, because R's question, in the therapeutic setting, is about whether to believe that she believes that p, and this question *is* approached by R 'in the first-person deliberative spirit', which is to say that this question is transparent to the question of whether she believes that p. R is thus *not* 'alienated from' her belief that she believes that p. Likewise, when the akratic gambler recognizes that she will choose to gamble, this may alienate her from her verdict that she should not gamble, but her question about whether to believe that she will choose to gamble is transparent to the question of whether she will choose to gamble. Thus, the latter question is raised from within Moran's first-person

⁴²⁹ See Allison 2004, pp. 344-5. I do not know how, if at all, this further standpoint distinction is meant to relate to Allison's three 'official' distinctions (as described in section 1 of this chapter).

⁴³⁰ Moran 2001, p. 3, p. 93.

⁴³¹ Moran 2001, p. 62.

deliberative perspective, and the gambler is not alienated from her predictive belief. Generally speaking, *whenever* one consciously takes up a question that one seeks to answer on the basis of reasoning, one exhibits a first-personal perspective and takes oneself to be governed by norms.

At one point Moran sums up the point of his standpoint distinction as follows:

The basic point can be expressed in a loosely Kantian style, although the idea is hardly unique to Kant. The stance from which a person speaks with any special authority about his belief or his action is not a stance of causal explanation but the stance of rational agency. In belief as in intentional action, the stance of the rational agent is the stance where reasons that justify are at issue, and hence the stance from which one declares the authority of reason over one's belief and action. (2001, p. 127)

As I said, there is indeed something in this picture that is congenial to Kant, namely, the idea that belief-formation is a rational activity that is subject to norms. *But for that very reason*, Moran's attempt to contrast a 'stance of rational agency' with a 'stance of causal explanation' is rather puzzling: the stance of causal explanation *is* a stance of rational agency because in trying to arrive at a belief about what caused some event one engages in an activity of belief-formation that is governed by the norms that apply to attempts to explain why things in nature happen as they do. *All* conscious belief-formation, and *all* reasoning aimed at belief-formation, is responsive to reasons and hence 'first-personal' in Moran's sense, including reasoning toward causal, explanatory beliefs. For Kant, whose standpoint distinction is partly but essentially meant to track the difference between two forms of reasoning, an appeal to a feature that is essential to *all* conscious reasoning (responsiveness to norms, first-personality) cannot afford us with a distinction between standpoints.⁴³² As I suggested earlier, for Kant the distinction is to be drawn by appeal to the difference between the distinctive subject matters of *norm-giving* reasoning on the one hand and norm-governed empirical reasoning on the other hand (see section 4 above). It seems, then, that the very concept of a standpoint is used differently by Kant and by Moran.⁴³³

This is not a criticism of Moran. There are, as Hannah Ginsborg reminded me, no analytic constraints on how to use the word 'standpoint'. Given his primary interest in the phenomenon of 'alienation', it is perfectly legitimate for Moran to draw a standpoint distinction that allows him to conceptualize this phenomenon in a systematic way. However, I briefly want to indicate that there is one substantive disagreement between Moran and Kant which might explain the different use they make of the concept of a standpoint. Moran presents his scenarios of the Freudian believer and the akratic gambler as if there was something special or defective about cases in which one considers one's mental states (beliefs or choices) as empirical *explananda* rather than as the result of one's being responsive to norms that govern the formation

⁴³² Thus, Allison's notion that "the idea of spontaneity is inseparable from the first-person standpoint" (1996, p. 127) is somewhat vacuous, because there is no third-person standpoint that is not also a first-person standpoint. The idea of spontaneity is inseparable from our self-awareness as subjects of reasoning from both Kant's empirical and normative standpoint (see section 4 of this chapter).

⁴³³ Moran's two standpoints can be adopted simultaneously, and in fact one *always* adopts the 'stance of rational agency', relative to some attitude, if one adopts any standpoint at all. Kant, by contrast, holds that if one occupies a standpoint, then one does so 'globally', not merely 'locally' or relative to some attitude. Adopting a standpoint for Kant implies that one does not and indeed cannot simultaneously occupy another standpoint. If one occupies the normative standpoint, then one tries to settle questions about what ought to happen and represents objects in ways that are unsuitable for the purpose of empirical cognition, and this rules out the possibility of occupying the empirical standpoint from which one tries to settle observational, explanatory or predictive questions about objects that cannot coincide with questions about how objects ought to behave (cf. section 5 of this chapter).

of the relevant states. Occupying a 'third-personal, non-deliberative perspective' from which one discovers, explains and predicts one's mental states is regarded as 'alienating'; and Moran thinks that our ordinary ('first-personal', 'deliberative') perspective toward our own mental states implies a kind of control that can be called upon to explain our *epistemic* authority with regards the question of what those states are.⁴³⁴ This control thus serves as the foundation of (what is special about) *self-knowledge*. By contrast, for Kant, every formation of mental states, *if* these are conceived as natural states of the *empirical self*, must be regarded as the necessary result of naturalistic processes; and there is nothing alienating about considering our own beliefs and choices in this manner.⁴³⁵ Correspondingly, Kant wholly rejects the suggestion that there is a perspective on ourselves that give us a special epistemic access to our own mental states. Our governance of our mental states through spontaneous responsiveness to norms and reasons cannot be invoked in an attempt to account for self-knowledge, because we can *know* ourselves only as phenomena, and because representing oneself in terms of the concept of spontaneity destroys the legitimacy of any claim to knowing oneself as phenomenon.

For Kant, the idea that self-knowledge can be disassociated from what can be empirically observed smacks of dogmatic, Cartesian Rationalism. This is a moot point: Moran takes himself to offer an account of self-knowledge that preserves the intuition that there is something special about self-knowledge without settling on the problematic Cartesian connotations (concerning the ability to directly and infallibly introspect what our mental states are) that this intuition is often taken to have. I cannot resolve this issue here⁴³⁶; the important point for the purposes of this

⁴³⁴ Moran 2001, p. 134: "A person is credited with first-person authority when we take the question of what he *does* *believe* to be settled by his decision as to what he *is* to *believe*."

⁴³⁵ Kant *does* think that there are special *methodological* difficulties that afflict the attempt to learn about one's own mental states (see my chapter II, section 2 and chapter VI, section 1). But this is an aspect of Kant's philosophy of science; it has nothing to do with what Moran calls alienation.

⁴³⁶ Kant and Moran are in agreement about the idea that *if* one's mental states are treated as items of empirical knowledge, then self-knowledge is exactly like any other kind of knowledge. From this, Kant infers that self-knowledge really is just like any other kind of knowledge, since he accepts that for some event or state to be an item of knowledge just is for it to be a potential object of empirical discovery. Since Moran does not make this inference, he must believe that an event can be an item of knowledge even if the source of knowing this event is independent of empirical discovery. By implying this Moran does not mean to commit himself to any dubious kind of rationalist access to non-empirical faculties. But it is hard to see how he can avoid such a commitment if he seriously wants to defend the view that we *know* our mental states non-empirically. It must be noted here that there is no need to disparage the intuition that 'transparency' is an important determinant of the belief states of a subject, if that subject is not afflicted by some kind of pathological condition to be examined by psycho-analysts. But it seems that this intuition can be shared by a third-person observer. If R seeks to determine whether Q believes that p, then the fact that p is the case can be one very relevant source of evidence for ascribing the belief that p to Q. How strongly this counts in favor of the ascription will depend on what p is and on Q's evidential situation. However, it seems clear that this cannot be the *only* kind of evidence that is relevant for the ascription: whether Q believes that p is a question that cannot be settled without considering episodes of Q's behavior, including his inferential behavior, and without considering Q's other mental states (beliefs, desires, and others). Let us refer to this evidential basis as *e*. Now, the reason why R must take into account *e* for determining whether Q believes that p is that the factors that comprise *e* determine whether it is in fact the case that Q believes that p. Whether or not Q believes that p depends in part on how he behaves and on what his other mental states are and how they are connected. Given this point, it follows that Q himself must, when determining whether or not he believes that p, take *e* into account. And that implies that he is really no better off than a 'third-person observer' in determining what his mental states are. I do not see what Moran's account does to challenge this point, unless the account somehow implies that each subject has, as far as her own mental states are concerned, either (a) an immediate introspective access to *e* that a third-person observer lacks or (b) an immediate introspective access to

appendix is that part of the reason why Kant's standpoint distinction does not appeal to the contrast between the first – and the third-personal perspective is that for Kant there is *nothing (epistemically) special* about first-personal deliberation. We can illustrate this point by reminding ourselves of Kant's claim that no person is in a position to know the precise content (and hence the moral quality) *of her own as well as any other person's maxims*. For Kant, 'practical knowledge', in the sense intended by Anscombe and, it seems, Moran (i.e., non-inferential, non-observational knowledge of one's own intentions), is impossible.

A final contrast from which Kant's standpoint distinction must be disassociated is that between a 'subjective' and an 'objective' perspective. Thomas Nagel and Peter Strawson suggest, each in their own way, that the perspective from which we think of ourselves and others as free and responsible is a *subjective* perspective that contrasts with the *objective* point of view we take when we explain the behavior of objects scientifically and by appeal to causal laws. Now, the concepts of subjectivity and objectivity are notoriously difficult to interpret, and a thinker's usage of these concepts is bound to reflect her overall metaphysical and epistemological commitments. For Kant, the representation of objects in terms of concepts or ideas such as freedom or spontaneity is, in one perfectly clear sense, *less subjective* than the representations of objects we come up with in contexts of causal, scientific explanations: the reason is that the former representations are *more intersubjective* ('more than human') than the latter because they are less dependent on the 'subjective', specifically human element in scientific cognition (spatiotemporality). The same is true with respect to the fundamental principles that govern our normative reasoning, such as the categorical imperative: these principles are valid for all rational beings. Kant does not deny that *within* the domain of cognition from the 'merely human' standpoint, we can draw various distinctions between merely subjective and objective representations.⁴³⁷ But generally speaking, Kant's conception of objectivity, subjectivity, and intersubjectivity cannot be shared by realists such as Nagel and Strawson because they treat space and time not just as merely human forms of sensibility. They portray time and space as absolutely real, as metaphysical conditions of being rather than as epistemic conditions of empirical cognition. Within such a framework – this is particularly clear in Nagel's case – any kind of conception that refers to the atemporality of the thinking subject and that does not answer to the standards of explanation imposed by the empirical sciences is to be regarded as confused and (in a disparaging sense) subjective.⁴³⁸

Hence, it is rather unhelpful to apply the objective-subjective distinction from contemporary discussions to Kant's contrast between two distinctive standpoints. But there is still room for interesting disagreements here. Nagel thinks that what he calls the 'standpoint of autonomy' is subjective and that the standpoint of naturalistic explanation is 'objective' because the former draws on ideas that do not withstand reflective scrutiny, whereas the latter discloses the real causes of our actions. In particular, he rejects as hopelessly confused the idea, employed from the standpoint of autonomy, that we are free and responsible agents that form our choices on the basis of reason. Kant, by contrast, thinks that our self-conception as free, responsible and

her mental states that renders it epistemically irrelevant. Both (a) and (b) exhibit assumptions that are at least reminiscent of 'Cartesianism'.

⁴³⁷ This allows for some interesting connections between Nagel, Strawson and Kant; for instance, all of them are quite prone to relegating colors, sounds, and smells to the domain of qualities that do not count as the properties of empirical objects in the strict and proper sense.

⁴³⁸ Nagel 1989, p. 238.

norm-governed agents is not a 'subjective' artifact of confusion dissolved by the discoveries of natural science, but rationally defensible on firm (albeit not on scientific) grounds. Nagel affirms whereas Kant rejects the idea that the natural sciences disclose the one paradigm of objectivity and rational defensibility.⁴³⁹ But most importantly, Kant would hold that someone who occupies the standpoint of natural science itself exhibits a kind of norm-governed, free agency that is incompatible with determination by natural events. So Kant's objection to Nagel's picture is that the standpoint of natural science cannot one-sidedly be contrasted with the standpoint of autonomy that Nagel takes to be compromised by scientific explanations: for Kant autonomy is itself a precondition of scientific objectivity. Hence, a most crucial difference between Kant and Nagel concerns their conception of what is required for *scientific* objectivity and rationality. It is this important disagreement, Kant's conception of *epistemic* freedom, rationality, and normativity, and the implications these issues have for Kant's defense of our belief in *practical* freedom that I will try to explain and defend in the final couple of chapters of my dissertation.

I want to conclude by indicating that the discussion in this appendix seems to me to show that the aim to thoroughly understand Kant's standpoint distinction has a significance that goes beyond the importance of philological accuracy. The attempt to correctly interpret Kant's talk about standpoints is not solely in the service of the philosophically dispassionate goal to carve out "the frozen object [of] 'Kant's philosophy itself'"⁴⁴⁰, but also guided by the goal to understand and conceptualize interesting and relevant philosophical disagreements about (as we saw) the nature of practical reasoning, self-knowledge, and objectivity.

⁴³⁹ This is a moot point: Nagel himself acknowledges one domain of objectivity which has nothing to do with scientific explanation. He acknowledges that there are practical norms of prudence and morality that can be shown to be objectively valid in a sense that is (somehow) analogous to the objective validity of scientific explanations. Now, while Nagel tries to defend the objectivity of these practical norms, he never retracts or qualifies his idea that our self-conception as autonomous, free and responsible agents is conclusively refuted by naturalistic explanations. In other words, Nagel tries to be a realist about certain norms and values without doubting that a naturalistic worldview is both correct and incompatible with our freedom and autonomy. Notice here the contrast with Kant: For Kant, the notion that practical norms have a kind of objective validity and authority stands and falls with the notion of autonomy, or with the notion that our reasoning toward these norms is not determined by the contingent natural states that make up our psychological constitution (see my chapter V). This is why for Kant the normative standpoint is a standpoint of autonomy: the standpoint from which we cognize objective normative principles that express rational necessities is a standpoint from which we must regard ourselves as purely spontaneous beings that can legislate practical laws. Nagel's attempt to disassociate the idea of rational defensibility of practical norms from the idea of freedom, and to drive a wedge between a standpoint from which we confusedly regard ourselves as free and an objective standpoint from which we recognize objectively valid normative principles would hence be unacceptable to Kant. As far as Nagel is concerned, the problem is how he wants to account for the idea that we recognize objective practical principles without an appeal to the freedom of practical reason or an appeal to some kind of obscure intuitionism.

⁴⁴⁰ Bennett 1984, p. 112.

Chapter IX: Freedom of Empirical Thought

In the preceding chapter, I argued that Kant's standpoint distinction implies that his notion of free, norm-governed agency is not limited to the exercise of our will but extends to the exercise of our ('higher') epistemic faculties in the formation of judgments about the empirical world. In this chapter, I examine the notion of freedom and agency that this picture requires.

1: Commentators on Kant on freedom of thought versus Kant on freedom of thought

We can begin by noticing that commentators tend to deny that Kant conceives of our theoretical attitudes – what we call beliefs, and what Kant calls 'assents' – as free and imputable. They suggest that Kant's notions of epistemic and practical spontaneity differ sharply, and that Kant considers beliefs *passive* states that we 'find ourselves with' as a result of causal necessitation:

But the freedom required by morality should not be confused with the 'spontaneity' characterizing the understanding. (Michael Friedman 1996, p. 438)

Kant...sides with Hume over Descartes on the issue of direct doxastic voluntarism...sufficient objective grounds typically not only license but also necessitate firm assent. In other words, once we acquire sufficient objective grounds for *p*, we typically just find our assent to *p* following along. (Andrew Chignell 2007a, p. 327)

[Perceptions] synthetically necessitate... the perceiver's assertoric belief in a corresponding propositional content. (Robert Hanna 2005, p. 264)

The I that thinks will be phenomenal and causally determined. (Patricia Kitcher 1993, p. 140)

Kant is leaving open the possibility that the being which thinks might be something 'which is not capable of imputation'. It might, in other words, be an *automaton spirituale* or *cogitans*, a thinking mechanism. (Wilfried Sellars 1970, p. 25)

I think that the conception of belief that is being attributed to Kant here rests (if only in part) on a very compelling philosophical intuition: beliefs, unlike choices, are not subject to our voluntary control. Believing is a matter of answering to the way the world is, whereas choosing is a matter of making the world answer to one's conception of how the world should be. This intuition seems to lend support to a picture on which belief is a passive phenomenon that merely happens to or in us, or according to which belief is an effect that occurs as a result of causal necessitation when our sense organs are affected with perceptual input.

But however compelling *we* may find this passive picture of belief, it must be acknowledged that there is no direct textual support for the idea that Kant himself holds such a picture. Rather, there are many passages that strongly suggest that Kant rejects the idea that the exercise of our cognitive faculties is causally necessitated (for instance, by perceptual input):

In lifeless, or merely animal nature, we find no ground for thinking that any faculty is conditioned otherwise than in a merely sensible manner. Man, however, who knows all the rest of nature solely through the senses, knows himself also through pure apperception: and this, indeed, in acts and inner determinations which he cannot regard as impressions of the senses. He is thus to himself, on the one hand phenomenon, and on the other hand, in respect of certain faculties the action of which cannot be ascribed to the receptivity of sensibility, a purely intelligible object. We entitle these faculties understanding and reason. (CPR, A 546/B 574)

He [Schulz, a fatalist who denied that we have free will]...presupposed: that the understanding has the capacity to determine its judgment according to objective reasons that are valid at any time, and does not stand under the mechanism of merely subjectively determining causes, which may change subsequently; he always assumed freedom of thought, without which there would be no reason. **In the same way**, he must presuppose freedom of will in acting.... (RezSchulz, 8: 14; my emphasis)

If an appearance is given us, we are still completely free as to how we want to judge things from it. The...appearance was based on the senses, but the judgment on the understanding.... (Prol, 4: 290).

These passages shows that Kant holds that we are free, even completely so, in the formation of our judgments about the empirical world. In the course of this chapter, I will argue that Kant has a coherent and interesting conception of freedom of empirical belief (or judgment) which respects the intuition that beliefs are not under our voluntary control, and which bears both striking similarities and dissimilarities to his conception of freedom of choice.

2: The subjective and the objective grounds of empirical judgment

Kant designates the faculty of understanding as the capacity to judge (CPR, A 69/B 94).⁴⁴¹ He repeatedly says that the exercise of this capacity depends on the way in which its *objective and subjective grounds* are brought together by the subject of thought. The idea that judgments of experience result from a unification of objective and subjective grounds is, I think, crucial to Kant's conception of freedom of empirical thought. So let us try to get clear about this idea.

Kant's most explicit discussion of the notions of the objective and subjective grounds of cognition occurs in a section of the Canon of the Critique of Pure Reason entitled, "Opining, Knowing, and Believing" (CPR, A 820/B 848-A 831/B 858). Kant here introduces a generic label for the kind of doxastic attitude that accompanies all judging: 'Fürwahrhalten', 'holding to be true', often translated as 'assent'. Kant says that judging and assent may or may not rest on *objective grounds* but in any case have to rest on *subjective grounds*. He adds that judgments and assents that rest on ('sufficient') objective grounds are intersubjectively valid and communicable, whereas judgments and assents that rest *solely* on subjective grounds lack these properties.

I think that Kant's discussion in the Canon suffers from a severe problem: he applies the notions of subjective and objective grounds indiscriminately to all classes of judgment.⁴⁴² I do

⁴⁴¹ Chignell 2007b, p. 35 claims that "strictly speaking", in Kant 'judgment' is a logical concept similar to the contemporary notion of a proposition; when Kant speaks of forming judgments, "what he really means is forming assents which have a...judgment as their object." There is no textual evidence for this claim. The very notion of a 'capacity to judge' implies that judging is *something we do*, by exercising this capacity. This is not to deny that Kant *also* considers judgments as logical objects similar to our notion of a proposition. For a very good account of the logical and psychological connotations of the notion of judgment in Kant, see Longuenesse 1998.

⁴⁴² This problem afflicts, for instance, Kant's idea that *both* the assent to an erroneous empirical judgment *and* the assent to the proposition that God exists (an instance of the species of assent Kant here calls *Glauben*, 'faith') rest on subjective grounds that lack objective sufficiency. But if the mark of objective sufficiency is intersubjective validity and communicability, the judgment that God exists is – in Kant's system – a paradigm case of a judgment that has objective sufficiency. *Every finite rational being* has rational (practical) grounds for the thought that God exists, and hence the judgment enjoys a kind of universality that does not attach to judgments that are our best candidates for objective *empirical* certainty, such as those of physics: while these latter judgments would be inaccessible to discursive thinkers that lack human forms of sensibility, we would be able to convey to such beings (and they to us) shared grounds for our faith that God exists. Chignell 2007a, p. 327 suggests that we should confine Kant's notion of 'objective grounds' to factors that can ground *empirical knowledge*, such as perceptual and memorial states as well as objectively grounded assents (to empirical propositions) that we already hold and that allow for inductive reasoning. This suggestion fails because Kant (in the Canon and elsewhere) explicitly applies the concept of objective grounds that are sufficient for knowledge to moral judgments (which, like the judgment that God exists, are universally communicable); so the notion of 'objective grounds that suffice for knowledge' cannot be defined in terms of empirical evidence. This is not the only problem that afflicts Kant's discussion at CPR, A 820/B 848-A 831/B 858. For instance, his notions of 'subjective sufficiency' and 'conviction' are very obscure and ambiguous. Chignell's attempts to clarify

not want to dwell on this problem here, for it is only empirical judgments that I will be concerned with in this paper. So let us see whether we can give a clear sense to the contrast between objective and subjective grounds of judgments that state how things are in nature.

Here a passage from the Dialectic is instructive (CPR, A 294-5/B 350-1; my emphasis):

For truth or illusion is not in the object, in so far as it is [sensibly] intuited, but in the judgment about it, in so far as it is thought. It is therefore correct to say that the senses do not err – not because they always judge rightly but because they do not judge at all. Truth and error, therefore, and consequently also illusion as leading to error, are only to be found in the judgment, i.e. only in the relation of the object to our understanding. In any knowledge which completely accords with the laws of the understanding, there is no error. In a representation of the senses – as containing no judgment whatsoever – there is also no error. (...) The [understanding] would not [of itself fall into error], since, if it **acts [handelt]** only according to its own laws, **the effect (the judgment)** must necessarily be in conformity with those laws; conformity with the laws of the understanding is the formal element in all truth. (...) Now since we have no sources of knowledge besides [understanding and sensibility], it follows that error is brought about solely by the **unobserved influence of sensibility on the understanding, through which it happens that the subjective grounds of the judgment enter into union with the objective grounds** and make these latter deviate from their true function....

Kant here considers empirical judgment *an effect* that results from an *act* of the understanding; an act, however, that is essentially constrained and affected by something other than the understanding, namely, sensibility. At CPR, A 295/B 351, Kant speaks of empirical judgment as a "composite effect" "of the [effects of] understanding and of the sensibility". And, as is clear from the above passage, Kant equates an empirical judgment's 'objective grounds' with the laws and concepts of the understanding and its 'subjective grounds' with intuitions or perceptions.⁴⁴³

To understand why Kant associates concepts with objectivity and the senses with subjectivity, we need to remind ourselves of some of the general tenets of Kant's epistemology. For Kant, the impressions that we receive through the senses do not of themselves represent objects (CPR, B 130-1; B 137-8; B 219; B 233-4; Prol 4: 290). Having a sensible impression or perception is, in itself, *subjective* because it is just a private episode in a given person's mind that cannot, as such, be shared with other subjects of perception (CPR, B 139-40; B 142). In order for a given perception to represent an object that is something distinct from the perception

Kant's position are instructive in many ways, but not (in my view) successful.

⁴⁴³ The relation between the terms 'intuition' and 'perception' in Kant is not straightforward. Arguably, perceptions are (unlike mere intuitions) conscious and conceptually structured insofar as they fall at least under the mathematical categories (quantity, quality). On the other hand, perceptions by themselves do *not* represent an intersubjective object of experience (CPR, B 219; Prol, 4: 298). The peculiar middle status perceptions have in Kant comes to light in his distinction between judgments of perception and judgments of experience in the Prolegomena (Prol, 4: 297 ff.). In this chapter, I am *only* concerned with judgments of experience. One may object to my account of empirical judgment formation that I am committing Kant to the notion that there is a wholly subjective, pre-conceptual *given* that is prior to and independent of the domain of judgments, concepts, and objectivity. Abela 2002 claims that the rejection of the idea of the sensible given is *the* essential point of Kant's critical philosophy and the Copernican revolution. While I find much of value in Abela's analysis, the idea that Kant's epistemology is exclusively informed by Sellarsian/McDowellian concerns about 'the given' strikes me as somewhat anachronistic. For the purposes of this chapter, I want to stress only that Kant clearly wants to endorse at least the following two points: (I) as epistemologists, we can and should – through abstraction – carefully separate the sensible (subjective) and the conceptual (objective) elements in our representations; (II) as subjects of empirical thought, we can and should become conscious of *sense-data*, such as colors and tastes, that are wholly subjective and that cannot (correctly) be ascribed to mind-independent phenomena.

itself (CPR, A 105) and that is cognitively accessible to other perceiving minds, it must be brought under concepts of two types: first, under pure concepts that enable the representation of a public, empirical object in general (the categories; CPR, A 111-2; B 165; A 93/B 126); second, under empirical concepts that determine the specific character of this object. So in the domain of empirical judgment, the contrast between the subjective and the objective grounds of cognition amounts to the contrast between the sensible given and the conceptual rules for combining perceptions in a manner that enables object-directed thought.

Now, Kant says that in a case of *error*, the subjective grounds of an empirical judgment 'enter into union' with its objective grounds. This is *prima facie* puzzling: for Kant, sensibility and concepts *must* 'enter into union' for empirical cognition to be possible – intuitions and conceptual representations are individually necessary and (only) jointly sufficient for the recognition of empirical objects (CPR, A 92-3/B 125). So it seems that Kant must consider the union between sensibility and concepts *both as a necessary enabler and as a potential defeater* of empirical cognition. Kant is aware of this ambiguity. Early on in the *Analytic*, he states:

The understanding can intuit nothing, the senses can think nothing. Only through their union can knowledge arise. But that is no reason for confounding the contribution of either with that of the other; rather it is a reason for carefully...distinguishing the one from the other. (CPR, A 51-2/B 75-6).

In the *Dialectic* (CPR, A 295/B 351) he says that when sensibility is "subordinated" to (the rules of) the understanding, "it is the source of real modes of knowledge". It becomes a source of error when "it influences the operation of the understanding, and determines it to make judgments". This is very ambiguous: does Kant mean only that the senses *influence* an erroneous judgment (by misleading the understanding) or that they are *causally responsible* for it? It seems that he must have the weaker point in mind: if the understanding was literally determined to err by sensibility, Kant's claim that the senses do not err would not have much substance. That he indeed endorses only the weaker point is clear from the *Prolegomena*:

[empirical illusion] is not ascribed to the senses, but to the understanding, whose lot alone it is to render an objective judgment from the appearance. [Note: what Hatfield translates as "is not ascribed to the senses" is "kommt nicht auf Rechnung der Sinne". This is important because the German word for imputability is 'Zurechnung', and hence Kant here considers, literally, the source of the imputability of an empirical error. This source is the understanding: "allein der Schein kommt nicht auf Rechnung der Sinne, sondern des Verstandes, dem es allein zukommt...ein objektives Urteil zu fällen."] (Prol, 4: 291)

The overall picture thus seems to be this: there is a legitimate unification of subjective (sensible) and objective (conceptual) grounds that results in empirical knowledge; this is to be contrasted with cases where sensible data get illegitimately confounded with conceptual representations, the result of such confounding being an empirical error that is to be imputed to *us* as subjects of thought. In the claim, "if an appearance is given us, we are...completely free as to how we want to judge things from it...the judgment [depends] on the understanding...", the 'we' (designating the subject of thought) is identified with the understanding qua source of empirical judgment.

To get a better grip on what such 'confounding' amounts to, we can take a clue from Kant's discussion of empirical error in the *Canon* (CPR, A 820/B 848). He says that such error – what he calls 'persuasion' – arises when a judgment has its ground only "in the special character of the subject" rather than "the character of the object". This terminology refers us to Kant's discussion of sensations such as *taste* that we get when we are affected by objects and that do not

indicate an objective quality: "The taste of a wine does not belong to the objective determinations of the wine...but to the special character of sense in the subject that tastes it" (CPR, A 28). I think we should read Kant as suggesting that in a case of empirical illusion, the subject of thought *carelessly* conjoins a merely subjective sensible impression with objective conceptual representations. Suppose I recognize an object as yoghurt; that is, I conceptualize it in terms of the valid empirical concept of a dairy product that has certain chemical and nutritional features. This recognition is, in turn, guided by a more fundamental concept that regulates my representation of an objective world: the concept of substance, which tells me to differentiate between a (at least relatively permanent) substratum that persists through changes and properties this substratum has for a certain period of time. Eating the yoghurt, I have an impression of sourness. My senses present the yoghurt *as* sour to me. So far there is no error, because I have not yet *assented* to this impression. The error occurs when I give in to my perception of the yoghurt as sour, when I *judge* that sourness is a quality that inheres in the yoghurt, in short: when I mingle some merely private element of my sensible apparatus with the objective grounds of cognition (empirical and pure concepts). Kant has another example:

And then it is not the fault of the appearances at all, if our cognition takes illusion for truth, that is, if intuition, through which an object is given to us, is taken for the concept of the object, or even for its existence, which only the understanding can think. The course of the planets is represented to us by the senses as now progressive, now retrogressive, and herein is neither falsehood nor truth, because as long as one grants that this is as yet only appearance, one still does not judge at all the objective quality of their motion. Since, however, if the understanding has not taken good care to prevent this subjective mode of representation from being taken for objective, a false judgment can easily arise, one therefore says: they appear to go backwards; but the illusion is not ascribed to the senses, but to the understanding, whose lot alone it is to render an objective judgment from the appearance. (Prol, 4: 290-1)

The judgment that the planets move backwards is not wholly dominated by merely subjective impressions. It *does* achieve a relation to an object and even a correct specification of the character of this object: it synthesizes given intuitions under the concept of substance and determines the substance under a valid empirical concept ('planet'). This is Kant's point when he says that "all error into which the human understanding can fall is only partial, and in every erroneous judgment there must lie some truth" (Log, 9: 54). Moreover, the perceptual representation of the planets as moving backwards is *itself* infused with objective conceptual elements which reflect the perceiving subject's state of empirical knowledge: it is only on this basis that the perception can present the planets as moving backward. This presentation is cognitively innocent. The illegitimate intermingling occurs when the subject *carelessly* assents to these perceptions and ascribes what is only an element of her contingent perceptual apparatus to public objects (the planets) whose character and existence is independent of this apparatus.

This gives us a first approximation to Kant's conception of freedom of thought: an empirical judgment is not forced upon us by perceptual input (or by the external objects that afford us with such input). Judgment (and assent) results from an imputable exercise of our rational faculties, whereas being affected with sensible data by outer objects is a passive event.⁴⁴⁴

⁴⁴⁴ One might wonder whether Kant's examples show that our beliefs are *always* imputable to us. E.J. Coffman has suggested the following counterexample: due to an environmental glitch, my senses present to me what's actually an oblong *as* a square. Naturally, I form the non-inferential belief that the indicated figure is square. Doesn't it seem as if, in this case, my false empirical belief doesn't result from carelessness and thus isn't imputable to me? I agree that it is far-fetched to think that in this case the belief results from carelessness. But

But Kant's idea of freedom of belief also relates to the role that *our internal psychology* plays in belief formation. Kant distinguishes between a transcendental, spontaneous (productive) and a merely empirical (reproductive) mode of operation of the *imagination* (CPR, B 152). The empirical imagination is a source of habitual, psychological patterns of associating representations (CPR, A 100; B 140), and empirical error may occur when "the faculty of judgment is misled by the influence of [the empirical] imagination"(CPR, A 295/B 352).⁴⁴⁵ To say that we may be misled by our imagination does not imply that the imagination is responsible for our cognitive mistakes. To take a trivial example: as a result of conditioning, I might find myself connecting the thought of lemmings with the thought of suicidal tendencies. I have no control over this associative pattern. But if I adopt the belief that lemmings as a matter of fact exhibit suicidal tendencies, this is not a mere causal upshot of psychological customs beyond my control. Rather, the belief is ascribed to me as a free thinker. In a different case, the empirical imagination might cause me to associate a certain type of event A with a different type of event B after I have seen these events being regularly conjoined. While I have no control over these associative patterns, my careless judgment that A and B are causally connected is to be attributed to me as a thinker rather than to the workings of my inner psychology over which I have no control. Kant is here in a substantive disagreement with Hume. While Hume also holds that having a sensible impression is not sufficient to lead us to form a belief, he claims that our beliefs about 'matters of fact' are caused by blind, non-rational processes that stand under the psychological laws of the imagination.⁴⁴⁶ Kant agrees with Hume that our minds are subject to the workings of the empirical imagination (CPR, B 140), but he does not accept that those workings determine us to make judgments. Empirical judgments are not forced upon us by our contingent psychology, although the understanding (that is, the noumenal subject of thought) may carelessly mistake psychological (subjective) patterns of association for objective relations.

In sum, we can say that for Kant, freedom of empirical belief consists, *negatively speaking*, in the *absence of* the causal determination of judgment by perceptual data (or the external objects causing such data) and by internal, sub-personal psychological processes.⁴⁴⁷

3: Freedom of thought without voluntarism

Now the question is whether we can consider this conception of freedom of belief also in a positive manner, as a *freedom to act* rather than merely (negatively) as a *freedom from* outer or inner determination. At this point we must take into account the intuition that, as we saw at the beginning, partly motivates the passive picture of belief: namely, the intuition that belief, unlike choice, by its essence cannot be a state whose formation is up to us or which we can freely control in any deeper sense. This idea is powerfully expressed by David Wiggins as follows:

this doesn't mean that the belief is not imputable to the subject. Even in a case like this, the belief results from the subject's taking her perceptual data as grounds for assenting to a certain empirical proposition, and this 'taking of' and the resulting assent is not beyond the thinker's rational control. A thinker can always doubt the evidential force of perceptual data, and refuse to 'consent' to what they present – she can consider the possibility that her perceptions are non-veridical even when they are in fact veridical. See McDowell 1994, p. 26.

⁴⁴⁵ This is not in tension with Kant's claim that it is the *influence of sensibility* under which erroneous judgments are formed: Kant thinks that the faculty of the imagination is part of the faculty of sensibility (CPR, B 151).

⁴⁴⁶ See Hume 2009 (the Treatise), Book I, Part 1, Section 4; I, 3, 6; I, 3, 8.

⁴⁴⁷ Starting out with such a 'negative' conception of theoretical freedom is recommended by how Kant proceeds in his discussion of practical freedom, which is first introduced, in the Critique of Pure Reason, as the negative concept of freedom from necessitation by sensuous impulses; see CPR, A 534-5/B 562-3.

It is up to me to make up my mind. But how far is this from the platitude that if it is my mind then only I *can* make it up, or the truistic interpretation of 'you must believe what you think true'? It does not seem to follow from any 'duty' I may have to do these things that I am exempted from thinking what *is* true, or that it is up to me *what* to think true. And it cannot be up to me if belief is to retain its connection with the world. And this connection must, from the nature of belief and its onus to match the world, be subjection. If belief does not retain this connection and subjection it could not be belief. For unless I am God...reality is most of it precisely what is *independent* of me and my will. Its arrangement is precisely what is *not* up to me.... From which it would seem to follow that if my beliefs are to relate to the world at all, I simply have to lay myself open to the world in order to let the phenomena put their print upon me. How otherwise can my beliefs even aim at a correct account of the world? If my state is one which seals itself off from the outside, this surely enfeebles its claim to be a state *of belief*. For there is subtracted everything which distinguishes belief from fantasy.... (1969, p. 143)

We can extract the following line of argument from this passage: (1) Belief is, by its very nature, a state that aims at truth, i.e., at depicting an arrangement of reality that is independent of my will. (2) Hence, if my attitudes were under the immediate control of my wishes concerning what reality should be like, they could no longer be regarded as beliefs (rather than fantasies). (3) Thus, my beliefs have to be independent of my will. (4) Thus, my beliefs have to be regarded as purely passive phenomena: they cannot be 'up to me' in any non-trivial sense, and their formation requires laying 'myself open to the world to let the phenomena put their print on me'.

The inference from (3) to (4) rests on the idea that agency and freedom as such must involve the will, choice or decision. It is illustrated by Bernard Williams' claim that the fact that "it is not the case that belief is connected with any decision to believe" implies "the picture offered by Hume of belief as a passive phenomenon, something that happens to us."⁴⁴⁸

Now, Kant agrees with the first three steps of this argument:

The will does not have any influence immediately on assent [Fürwahrhalten]; this would be quite absurd. (...) If the will had an immediate influence on our conviction concerning what we wish, we would constantly form for ourselves chimeras of a happy condition, and always hold them to be true, too. But the will cannot struggle against convincing proofs of truths that are contrary to its wishes and inclinations. (Log, 9: 73-4. See also CPR, A 93/B 125.)

The line of thought in this passage could be reconstructed in terms of (1)-(3) just as Wiggins's argument, 'wish' corresponding to 'fantasy' and 'assent' to belief.⁴⁴⁹ But Kant does not move from

⁴⁴⁸ Williams 1973, pp. 147-8. He thinks that Hume goes wrong in supposing that it is a contingent fact about our minds that we cannot believe at will. The inference from (3) to (4) seems also, if implicitly, in play in Chignell's idea (see p. 195 above) that *since* Kant rejects doxastic voluntarism, he holds that our beliefs are passive states that simply 'follow along' the acquisition of perceptual data.

⁴⁴⁹ Bennett 1990 challenges this argument by imagining a creature, the Credamite, who can will to have a desired belief and then immediately forgets how she acquired it. But all this shows is that the Credamite can will *something* that has, as its effect, a belief-state. Nobody doubts that this is possible: for the Credamite, this 'something' is a strange psychological act of self-manipulation, whereas for us the 'something' is taking drugs or undergoing hypnosis. Bennett has not shown that a thinker can adopt a belief on the basis of volition while being fully aware of the (doxastic) irrationality of the belief; he must stipulate that for the Credamite the adoption of a belief that results from something she chooses is simultaneous with falling into oblivion about the origin and the doxastic credentials of the belief. Now, Bennett says that beliefs "are theoretically describable in statements about what would be the case about the believer if such and such came to be the case. A belief is a function from desires to actions". This is the crux of the issue: Kant can agree that *if we think of belief in those terms*, namely, as a state defined by its causal role, then the only constraints on coming to be in this state are causal factors. His anti-voluntarism relates to belief considered *normatively*, by appeal to the distinctively theoretical norms to

(3) to (4): that is, he does not infer from the fact that beliefs are not under the immediate control of the will to the claim that beliefs cannot be up to the subject of judgment in any non-trivial sense, or that beliefs must be considered wholly passive states. As we already saw, he accepts that empirical judgments result from *free, imputable acts* of the understanding. Since he denies that this freedom involves the will, the challenge is to specify a non-volitional sense in which we are free and active with regard to our cognitive responses to passively received perceptions.

Several contemporary philosophers who are friendly to the idea of epistemic agency or freedom suggest that forming a belief involves the use of our rational capacities, or a responsiveness to reasons:

...thinking is necessarily associated with reasoning – thinking guided by reasons – and reasoning cannot in general be a mere ‘going on’. In making an inference, a being is *ipso facto* an agent. (Tyler Burge 1998, p. 251). Judging, making up our minds what to think, is something for which we are...responsible – something we freely do, as opposed to something that merely happens in our lives. (...) This freedom, exemplified in responsible acts of judging, is essentially a matter of being answerable to criticism in the light of rationally relevant considerations. So the realm of freedom, at least the realm of freedom of judging, can be identified with the space of reasons. (John McDowell 1998a, p. 434).

But Wiggins and Williams do not deny that the formation of belief is guided by reasons. They would say that being guided by theoretical reasons reduces to awareness of the perceptual evidence, which causes belief. As David Owens puts it, "if anything is in control of the belief-forming process here, it looks to be the...experiences that motivate belief rather than the subject who enjoys these...experiences."⁴⁵⁰

I think that the Kantian point here is that the responsiveness to reasons that issues in judgment and assent has a *reflective* component: the subject of thought forms her belief on the basis of a representation of what her theoretical reasons are.⁴⁵¹ The thinker's reflective conception of what she ought to believe controls her beliefs: that is, her beliefs about the world result from being governed by, or *subjected to*, her conception of the quality of her evidence. It is by virtue of my endorsement of some perceptual datum as a sufficient reason for a belief that p that I *bring it about* that I believe that p; it is by virtue of my doubt about whether a perceptual datum supports the belief that I *bring about* a suspension of judgment; and it is by virtue of my verdict that a given perception which represents p as being the case is wholly subjective that I *bring about* a belief that not-p. Thus, a subject exercises a form of agency over belief-formation by virtue of controlling her epistemic attitudes through her reflection about what her evidence is and what one should believe (i.e., what beliefs are required, or permitted, or prohibited) in the light of this evidence. This form of agency may be called 'reflective control'.⁴⁵² I shall argue that the

which it has to answer. For a somewhat similar, more detailed response to Bennett, see Hieronoyimi 2006.

⁴⁵⁰ Owens 2000, p. 3.

⁴⁵¹ The reflective nature of the act of judgment must be distinguished from a more fundamental sense in which cognition involves a (self-)reflective element: namely, from the conscious combination of given intuitions under categories in what Kant calls the original synthesis of given intuitions. This kind of synthesis is more fundamental than the synthesis of empirical concepts in judgment, because it achieves a relation of given intuitions to objects which is a cognitive precondition for further determining this object through empirical concepts (it "precedes a priori all my determinate thought"; CPR, B 134). I am here concerned only with this further determination, in acts of conceptually determinate empirical judgment (but see my chapter X). It is a weakness of Pippin's 1987 (otherwise illuminating and pioneering) discussion of Kant's conception of theoretical spontaneity that he does not distinguish between these two dimensions of (self-)reflective cognition.

⁴⁵² The idea of reflective control is explicit or implicit in Burge 1998, McDowell 1998, Moran 2001 and Korsgaard

proposal to understand freedom of belief in terms of reflective control is coherent and non-trivial. For this purpose, I shall consider four objections that can be raised against this proposal.

First, one might hold that there is *no need or place* for such control: it couldn't do anything for us that our non-reflective awareness of the evidence (i.e., the worldly facts) isn't already doing; hence, judgments about what our evidence calls for, or what we ought to believe in the light of our evidence, look like "an idle wheel in our motivational economy."⁴⁵³ Now, it is certainly implausible to hold that mere perceptual awareness is sufficient to move us to believe one thing rather than another about the world. For example, unless I represent my perception of a certain sound pattern as a warning cry, and unless I draw on my belief that cries of this sort are characteristically uttered by jaybirds, I will not form the belief that a jaybird is nearby. The opponent of reflective control may admit this and insist that nothing in this picture requires attributing to me the notion that the perception of the warning cry constitutes a reason to believe that there is a jaybird nearby. But the conjunction of the perception of a sound pattern and of the belief that sounds like this are characteristically uttered by jaybirds does not automatically induce the belief that there is a jaybird nearby. Suppose that I have other beliefs that make it harder to interpret my perception: for instance, suppose I believe that there is another species whose exemplars utter sounds similar to those uttered by jaybirds, but I also believe that it is unlikely that this other species exists in my current area. Here my belief that there is a jaybird nearby results from *putting together* a number of evidential considerations, not merely by simply adding one to the other but by taking the conjunction of all these beliefs to license, or rationalize, the formation of the belief that there is a jaybird nearby. Or, suppose that I doubt the evidential force of my perception. For instance, I might think that given my proven incapacity to correctly interpret bird sounds, the fact that I associate a sound with crows is not a good reason to believe that the sound really comes from a crow. Here I accept that my perceptual input evidentially underdetermines what I should believe, and this controls my suspension of a belief about what caused my perception. But this suggests that in the original example, my adoption of the belief that there is a crow nearby depends on the fact that I (however implicitly) take my perception to constitute a sufficient ground for adopting this belief, i.e., that I do not doubt or question its evidential weight. The Kantian idea is, then, that our regulation of our doxastic attitudes typically depends on our reflective conception of what our theoretical reasons are.⁴⁵⁴

1996, pp. 92-3. As Owens 2000 points out, many of Kant's predecessors, such as Descartes or Locke (but not Hume) also held that we exercise this control over our beliefs. I want to anticipate three ways in which Kant's position here is distinctive. First, for Kant our reflective control over our beliefs is systematically tied up with our rational imperfection; second, our having such control stands and falls with our freedom from natural causality; third, Kant – unlike Burge and Moran – does not think that the control we have over our beliefs affords us with self-knowledge. I explain the first and the second of these points in the course of this chapter.

⁴⁵³ Owens 2000, p. 18.

⁴⁵⁴ Owens has a further objection to this Kantian idea: he claims that a reflective awareness of evidential considerations could never motivate an empirical belief because such considerations are never sufficient for assenting to empirical propositions. The crucial move in his argument is the introduction of an unanalyzed notion of conclusive evidence (p. 25) which, he thinks, is never actually instantiated: "purely evidential considerations underdetermine what we ought to believe until they receive pragmatic supplementation" (p. 26); "we can never have conclusive evidence about matters of empirical fact" (p. 27). Owens might just mean here that we lack *deductive* or entailing evidence for our empirical beliefs; but why should all conclusive reasons have to be deductive? Even if we bracket this issue and consider cases where our evidence *does* underdetermine what we ought to believe, this causes no problem for the Kantian idea of reflective control: Kant himself would say that in such cases, the subject ought either to suspend belief or, if she has some inconclusive evidence that tells in favor

This brings me to the second objection to the idea that we exercise freedom qua reflective control over our beliefs. The Kantian picture might seem to imply that we cannot form a belief *without going through actual episodes of reflection* by means of which we form for ourselves a conception of what beliefs are required or prohibited by our evidential situation; and this would yield a misconception of our ordinary, everyday patterns of belief-formation.⁴⁵⁵

Kant's view about the relation between reflection and judgment-formation is as follows:

Every prejudice is to be considered as a principle of erroneous judgment, and...erroneous judgments arise from prejudices. (...) The reason of this illusion is to be looked for in subjective grounds' being falsely held objective ones, **from a want of reflection that must precede all judging.** (...) **...we cannot and must not judge anything without reflecting,** that is to say, without comparing a cognition with the faculty (the sensibility, or the understanding) from which it arises. If we assume judgments without this reflection...prejudices, or principles for judging from subjective reasons that are falsely held for objective ones, arise. (Log, 9: 76. Compare Log, 9: 73)

This shows that Kant presents it as a *norm* of empirical cognition that we ought not to ("must not") judge without reflecting whether a passively received perceptual representation is merely

of p, adopt a correspondingly weak form of assent (in Kant's terminology, her assent should take the form of mere 'opinion' rather than 'conviction'). Now, one important disagreement between Kant and Owens, which I cannot settle here, is that Owens holds that epistemic justification depends on pragmatic factors, i.e., considerations such as: how important is the matter, how much time do I have for making up my mind, etc. Since he rejects doxastic voluntarism, he also rejects the idea that we can adopt beliefs on the basis of our reflective awareness of pragmatic considerations. Kant, by contrast, is an 'intellectualist' who holds that epistemic justification is determined entirely by truth-indicative factors. Notice that a Kantian intellectualist need not deny that pragmatic considerations can move us to (say) stop weighing the evidence for and against a certain proposition and to act on the basis of assuming ('opining') that a certain proposition, which is to some extent supported by the evidence, is correct: but none of this tells against the evidentialist-intellectualist idea that the subjective strength of our doxastic attitude should reflect the objective strength of the evidence.

⁴⁵⁵ I think that it is this worry, together with an externalist intuition concerning epistemic justification, that leads Chignell to attribute to Kant both a passive picture of belief-formation (see p. 195 above) and a corresponding externalist conception of epistemic justification: "For example, Joe's assent that the sky is blue can *be* justified simply by way of his having the experience of the sky as blue; he doesn't also need actively to introspect and determine *that* his perceptual experience is a sufficient objective ground for his assent. Kant would presumably agree with this" (2007b, p. 46). I do not think that Kant would agree. First of all, Joe's assent that the sky is blue would be considered by Kant – at least in the First Critique – a paradigm case of an assent that is *not only unjustified but false* and which results from the kind of careless intermingling of subjective and objective grounds that I discussed in section 2: "Colors are not properties of...bodies.... (...) They are connected with the appearances only as effects accidentally added by the particular constitution of the sense organs" (CPR, A 29). More generally, the idea that on Kant's picture the representations we passively 'acquire' through the senses constitute objective grounds that 'necessitate' our assents is doubly mistaken: first, it ignores the (as we saw) numerous passages in which Kant insists that the senses are not responsible for our assents; second, it ignores the fact that Kant holds that what is given to us (what we passively 'acquire') through the senses are *subjective* grounds of cognition whose objective evidential significance the subject of thought needs to determine reflectively. An even more extreme version of the idea that sensible intuitions necessitate belief is offered by Hanna, who claims (2005, p. 264) that for Kant even *non-veridical* appearances "synthetically necessitate... the perceiver's assertoric belief in a corresponding propositional content". But this flies in the face of Kant's claim that "when an appearance is given us, we are still completely free as to how we want to judge things from it", which is advanced *precisely* in a context where he is concerned with non-veridical appearances. Hanna also claims (*ibid.*, p. 263) that cases of perceptual illusion "imply the evidential force of some nonconceptual contents...for Kant false perceptual judgments based on illusions are sub-rationally non-inferentially warranted". But for Kant, a judgment based on perceptual illusion is clearly not warranted in *any* sense.

part of our subjective sensibility (i.e., is merely "accidentally added by the particular constitution of the sense organs"; CPR, A 29) or whether it can be reflected under valid empirical concepts that determine the empirical constitution of public empirical objects. But he does not hold that our judgments always *are* preceded by actual reflection, for he is well aware that we do not actually comply with this normative demand on a typical basis.

Notice that it does not follow from the fact that our empirical judgments often do not follow from episodes of explicit reflection that they are not under our reflective control in those cases where such reflection is absent. Consider the analogous issue in the practical domain: our specific choices can follow from our conception of what we have reason to choose without being preceded by specific episodes of reflection, or even without being accompanied by an explicit representation of what is worth doing, because we typically choose on the basis of *general* principles of willing ('maxims'). These principles embody chosen, standing commitments to accept certain things rather than others as grounds of choice and thereby foreclose our answers to a variety of specific practical questions. (My choice to get dressed before I leave the house is surely under my reflective control, but very rarely preceded by explicit reflection.) Likewise, our cognitive responses to perceptual input can be based on standing commitments to treat certain types of perceptions as grounds for adopting certain types of beliefs; commitments that foreclose our answers to a variety of theoretical questions.⁴⁵⁶ This seems especially plausible since we are faced with *recurring* patterns of perceptual input. The above passage shows that Kant acknowledges a theoretical analogue of general principles of choice: he posits freely adopted "principles for judging" that attribute evidential significance to certain recurring perceptual patterns and that control the 'automatic' formation of specific beliefs. In both the practical and the theoretical case, the idea that a person's unreflective choices and beliefs are typically under her reflective control can be supported by the fact that typically a person is at no loss – when pressed, or when in a reflective mood – to make explicit the reasons that implicitly guide her formation of choices and beliefs. Correspondingly, cases where people cannot come up with reasons for beliefs that they hold are typically cases where their beliefs are not under their reflective control – cases in which they are alienated from their doxastic states.⁴⁵⁷

Now, the principles for judging that Kant mentions in the case of belief-formation are *prejudices* that grant objective evidential weight to perceptual patterns that are in fact wholly subjective. So, the adoption of a disposition to accept certain patterns of perceptual input as grounds for adopting beliefs without conscious reflection is, according to Kant, an epistemic vice.⁴⁵⁸ Why does Kant think that we ought to form our judgments on the basis of episodes of explicit reflection? My argument in section 4 below will shed light on this issue.

A third objection to the model of reflective control comes in the form of a dilemma. The view I am attributing to Kant holds that a thinker's doxastic attitudes are formed on the basis of

⁴⁵⁶ We should not confuse the idea that reflective control takes place implicitly, in the background of thought, with the weaker idea that it is always possible for a subject of thought to exercise such control even though typically this kind of control is not at all exercised. This weaker position seems to be in play in McDowell 1998a, p. 334.

⁴⁵⁷ The phenomenon of such alienation is explored in great depth in Moran 2001. A good example would be the case of someone who discovers, in psychotherapy, that she believes that was abused by her parents.

⁴⁵⁸ Kant considers the adoption of general principles that lead to *suspension of judgment* an epistemic virtue; but this does not mitigate the normative demand that we always ought to base our beliefs upon reflection, because a suspension of judgment is (unlike judging out of freely adopted habit) necessarily based on one's appreciation of the fact that the relevant judgment is not supported by conclusive reasons (cf. Log, 9: 74-5).

her (however implicit) representation of what her theoretical reasons are. The question is: must the subject's representation of what her theoretical reasons are *itself* be based on a representation of what the subject's theoretical reasons are for *it*? If so, then we seem to face the threat of a regress; but if the relevant 'higher-order' requirement *doesn't* apply to such 'higher-order' representations, why think it even applies to first-order empirical beliefs?⁴⁵⁹ To illustrate my response, I want to go back to our earlier example: suppose, again, that I perceive a certain sound pattern that I associate both with birds of type x and birds of type y, and that I believe that birds of type y do not live in my current area. To the extent that we consider me as a rational thinker, we must suppose that it is essential to my moving from my perceptual and doxastic states to the formation of the belief that there is type-x bird nearby that I take the former states to rationally motivate (i.e., to rationalize) the latter belief. The question, 'why believe that there is a type-x bird nearby?' makes perfect sense in first-personal deliberation or in an exchange with another observer, and certain ways of answering it reflect better on me qua rational thinker than others (suppose I adopt the belief in part on the basis of my mistaken belief that only type-x birds can utter this kind of sound). But notice that this why-question seems artificial and unmotivated with respect to my idea that the conjunction of my perceptual and doxastic states justifies the belief that there is a type-x bird nearby. The representation, 'my perception of a sound that might come from birds of type x and y, together with my belief that type-y birds do not live here, provides grounds for believing that there is a type-x bird nearby' (call this representation r) does not depend on a *further* representation that provides grounds for it (for r).⁴⁶⁰

This suggests that there is a need to attribute to a rational thinker a conception of what her theoretical reasons are that arises with respect to her empirical beliefs, but not (necessarily) with respect to her conception of her theoretical reasons for empirical belief. As rational thinkers who partake in the space of reasons, we take ourselves (and others) to believe things about the world for (good or bad) reasons, so there is a clear route from the formation of an empirical belief to a (more or less explicit, maybe wholly implicit) representation of what justifies this belief. But it is not part of our self-conception as rational thinkers that our representation of our reasons for believing things about the world must be based on further (representations of) reasons.⁴⁶¹

I now want to address a fourth and final objection to the claim that the idea of reflective control gives us an interesting notion of freedom of belief. Recall the original worry that it is not up to us what to believe since beliefs essentially aim at the truth. The idea that our beliefs are controlled by our reflective conception of what we ought to believe was meant to give us a non-trivial sense in which our beliefs *are* up to us. Now, one might argue that since our reflective conception of our theoretical reasons is *also* aimed at the truth, it is not up to us in any interesting

⁴⁵⁹ This dilemma has been suggested to me by E.J. Coffman. See Alston 1989 for worries along those lines.

⁴⁶⁰ Here I am presupposing the untenability of skepticism about perceptual knowledge. If such skepticism *cannot* be defused in one or another way, then the general idea that certain perceptions (maybe in conjunction with other empirical beliefs) provide reasons for adopting empirical beliefs cannot be considered 'self-sufficient', and the question, 'why do you take your perception of an x-bird sound to justify your belief that there is an x-bird nearby?' is not artificial or unmotivated (it draws attention to the possibility that one's perception may be part of a dream, or be caused by a vat, evil demon, etc.).

⁴⁶¹ This is not to rule out that if one represents certain perceptions/beliefs as constituting reasons for other beliefs, this representation *may* itself be based on further reasons. We can imagine evidential situations that are vastly more complex than the simple example that I have been using. In such situations, our representation of reasons for taking a certain perception to tell in favor of a certain belief may itself be based on further justificatory thoughts about what makes those reasons (good) reasons.

sense either, and thus the idea of reflective control cannot ground a non-trivial notion of freedom of belief. This objection assumes that we exercise free control over an attitude *only if* it is up to us to change our verdict about whether this attitude is justified or not. And I think that this condition is unacceptably strong. We can see this by pointing out that if it *were* a strict condition on free control that we may change our conception of how our attitudes ought to be, then many of our choices would not be within our free control. As *rational choosers*, it is in no way up to us to reject the principle that one must not lie for the sake of personal gratification. And yet, if my choice to tell the truth is governed by this principle, this looks like a paradigmatic case of free agency. *By analogy*, as *rational cognizers* it is in no way up to us to reject the principle that we ought to suspend a judgment that p if we have grounds for thinking that our perceptual awareness of p might be misleading. Consider a case where a sound seems like a crow's call to me but I am notoriously bad at interpreting bird sounds. If my suspension of belief about what caused the sound is governed by the epistemic principle just mentioned, then my withholding of assent is free. Or, to bring up an entirely different example that Kant would consider very important: as rational cognizers it is in no way up to us to reject the principle that one ought to believe in causal necessity. So, if my belief that the pen *must* fall if dropped is governed by this principle, then my belief is free. I would lack such freedom in the Humean scenario where my belief that the pen must fall is produced by sub-rational processes and occurs completely regardless of whether I consider beliefs in causal necessity rationally justified.

I want to summarize what I take to be Kant's response to Wiggins and Williams. Kant agrees with Wiggins that the formation of a belief requires that we lay ourselves open to the world to let the phenomena put their print upon us. But Kant denies that the passive reception of sensory data is sufficient for the formation of belief. On Kant's view, there is what I want to call a reflective gap between taking in perceptual data and judging what these data reveal about the world. Kant agrees with Wiggins that the will is not involved in filling this gap: our cognitive responses to perceptual input are not based on our sense of what our practical reasons are. But for Kant the rejection of doxastic voluntarism does not imply that beliefs are passive phenomena that merely happen to us. On Kant's view, our cognitive attitudes are free because they stand under the direct control of our faculty of understanding and its verdicts about what our theoretical reasons are. Forming such a verdict is not done for us by the passive reception of sensory impressions; it is something we must do as free thinkers. So, the idea that we exercise freedom qua reflective control over our beliefs is coherent since it does not grant the will immediate control over belief and hence does not blur the distinction between belief and wish or fantasy. This idea is also non-trivial, because we can informatively contrast it with the naturalistic Humean picture of belief against which Kant is reacting. Hume would agree with Kant that having sense impressions is insufficient for believing. But for Hume, the gap between the reception of sensory input and the formation of belief is not closed by reflection but by sub-rational psychological processes. In the Humean scenario, our conception of what beliefs would be rationally justified is wholly inefficacious.

I conclude that the notion of reflective control allows us to further characterize Kant's idea of freedom of thought: such freedom not only consists, negatively speaking, in the absence of determination by sensible causes, but also involves, positively speaking, the capacity to form doxastic attitudes (such as assent, or suspension of judgment) through our reflective appreciation

of objective theoretical norms and reasons.⁴⁶²

If we have a capacity to form doxastic attitudes whose exercise is irreducible to the reception of sensible data or to the occurrence of sub-personal psychological processes, we have a capacity for epistemic *agency*. One may insist that *only* executing one's intention through bodily movements can be considered a free action; but this brings to mind Hampshire's criticism that "it would be a crude metaphysics that implied that an action is necessarily a physical movement."⁴⁶³ Unless one can show that there are certain features of bodily movement that are essential to our conception of action⁴⁶⁴, it seems fair to conclude that the concept of free action is too indeterminate to prohibit us from applying it to the mental realm.⁴⁶⁵ For Kant, this concept is broad enough to cover the exercise of our rational capacities for choice and judgment.⁴⁶⁶

⁴⁶² One might worry about the idea that we are generally free/active with respect to attitudes that arise upon our recognition of what the evidence dictates. For instance, suppose I have promised to pick up my partner from the train station, but I recognize that because of my laziness I will not be able to make it in time. Here it may be said that my attitude of regret toward the proposition that I will not be able to keep my promise simply forms passively in me, as opposed to being actively formed by me. I think that Kant would say *I am* free and active in my formation of regret here, insofar as my attitude of regret is responsive to the objective moral law that prescribes keeping one's promises; the formation of such regret is hence as much an expression of my free moral self as the formation of the moral principle itself. One might object that surely there are other cases where emotional states that are *not* governed by objective practical norms passively 'pop up' in us when we realize the facts; for instance, I might feel disappointed upon recognizing that I am not brilliant. But conceding that this is possible does not in the least tell against the notion that we are free and active with regard to the doxastic states we form in response to our conception of what the evidence dictates. The difference between the emotion of disappointment and my belief is precisely that the latter is, whereas the former is not, fully within my rational control, namely, controlled by the exercise of my capacity to respond to objective reasons. Even though my feeling of disappointment pops up as a result of my epistemic achievement of recognizing my intellectual limits, the feeling is not (*unlike* the belief about my intellectual limits) governed by norms, but rather the upshot of my contingent psychology. I would not be susceptible to any form of criticism if I lacked this feeling.

⁴⁶³ Hampshire 1971, p. 91. See, however, the following footnote.

⁴⁶⁴ Hampshire 1971 goes on to suggest that there are reasons to reserve the notion of action to physical activity. He argues that actions are essentially (i) done at will and (ii) make "some recognisable change in the world" (p. 154). But why should we accept this definition? A crucial point for Hampshire seems to be (pp. 160-4) that mental actions are not publicly perceivable, and hence enjoy less of a (as he calls it) "definite reality". But it seems to me to be a very "crude metaphysics" that measures the reality of an act in terms of its perceivability.

⁴⁶⁵ Hieronymi 2009 argues that we should distinguish between *two kinds of agency*: first, what she calls 'evaluative control', a kind of agency that we exercise over our mental states through settling certain questions on the basis of reasoning; and second, the kind of 'manipulative' control that we exercise both over external objects and our mental states by making them answer to our purposes (e.g., one opens the fridge to get a water bottle, or one takes a pill to lose the belief that p). She considers the second case the *paradigmatic* kind of agency and control, and she argues that what is distinctive of it is that it involves both the will and 'reflective awareness'. I am sympathetic to the proposal to define a paradigmatic notion of agency that essentially involves the will, but reflective awareness does not seem to help this cause. Reflective awareness *can be* a crucial feature of belief-formation, and it *can also be* completely absent from exercises of manipulative control: for instance, I may open the fridge to get a bottle of water while being immersed in thoughts about the soccer match, without explicitly representing to myself the purpose of my action. Patterns like these seem to me to abound in everyday behavior.

⁴⁶⁶ Kant's concept of 'action' is so broad that it encompasses even the causally determined behavior of mechanical bodies (CPR, A 204/B 250). With regard to *free* agency, some commentators suppose that for Kant *only* free will makes us rational agents; see Allison 1990, p. 63, p. 221-2, p. 228 and Sellars 1970. This is related to the conviction that practical and theoretical spontaneity are wholly different notions (Friedman 1996, p. 438; Kitcher 1984, p. 122). Now, Kant indeed has *a* concept of 'action' related to free will that has no analogue in the epistemic case. When Kant says that the 'ought' of practical reason expresses "a possible action" (CPR, A 547-8/B 575-6) and that practical reason expects effects from its 'ought' judgments in experience, his concept of

4: Freedom of thought and alternative possibilities

I argued that the positive core of Kant's conception of freedom of thought is his appeal to our capacity to form beliefs by responding to objective reasons and norms. This raises an important question. Is our freedom of thought to be identified with the capacity to form our attitudes in accordance with the correct norms (so that we are only ever free to get things right; this is what Philipp Pettit and Michael Smith call 'freedom as orthonomy')?⁴⁶⁷ Or does our freedom of thought give us the option to get things either right or wrong? As I argued in chapter VI, Kant adopts this second model in the *practical* domain: he holds that it is characteristic of human freedom of choice that our ability to choose in accordance with the right reasons has what I called a shadow, namely, our pervasive propensity to *knowingly* choose incorrectly. But this idea seems inapplicable in the theoretical case. Anyone, like Kant, who rejects doxastic voluntarism cannot allow that we may knowingly believe the wrong thing.

Nonetheless, I want to argue that it is an important feature of Kant's conception of freedom of thought that it leaves us with alternative possibilities, and in particular with the possibility to form our doxastic states improperly. Belief, as such, aims at the truth: this provides a constraint on what may count as conscious formation of belief. But notions such as 'truth' or 'evidence' are very thin, even empty. Aiming at the truth allows one to partake in the game of belief-formation, so to speak; but the concepts of truth and evidence by themselves provide no concrete guidance as to what one should believe.⁴⁶⁸ To adopt a belief, a subject of thought must form for herself a representation of what counts as evidence, or of what considerations tell for or against the truth of a certain proposition. Hence, the notions of 'truth' or 'evidence' not only *allow* but indeed *call* for interpretation on part of the thinking subject. If it is always possible for us to

action relates only to an *outer* action, namely, to the production of an object of morally pure intention, such as the *actual payment* of a debt out of respect for the moral law. But this cannot be the only relevant concept of free action for Kant, for he thinks that it is *only at the inner level of choice* that complying with the moral law is always within our control (e.g., in attempting to pay my debt, I may have a car accident); see KpV, 5: 36-7 (see MM, 6: 214 for Kant's distinction between 'inner' and 'outer' actions). Likewise, we should take his reference to "actions of the understanding" (CPR, A 330/B 387) at face value, as designating a genuine kind of inner, mental agency. For further discussion, see chapter III, section 3; chapter VII, section 3; chapter VIII, section 3.

⁴⁶⁷ See Pettit/Smith 2003. A similar conception of freedom is in play in Wolf 1980, 1990, 2003.

⁴⁶⁸ Shah suggests (2003, p. 470) that "a competent user of the concept of belief must accept the prescription to believe that p only if p is true for any activity that he conceives of as belief-formation... when an agent raises the question of whether to believe that p, then, insofar as he is a competent user of the concept of 'belief', his grasp of the concept of belief will direct him to believe that p only if p is true." This attempt to represent 'truth' as a *prescriptive norm that governs* belief-formation strikes me as problematic. First, it is unclear why someone who rejects that one should believe that p iff p is true fails to possess the concept of belief; there are many handles one can get on the concept of belief, for instance, via the connection between belief and behavior. Second, if truth were to function as a prescriptive norm, it would have to provide guidance for doxastic deliberation, and it is a hard to see how this is supposed to work. It might be said that the norm of truth provides guidance in that it requires one to deem only evidential considerations relevant in the conscious steps preceding belief-formation. But, on Shah's constitutive model, if one discounts evidential considerations, then one ceases to be engaged in belief formation and ends up (say) daydreaming; and choosing to engage in daydreaming rather than belief-formation cannot itself be understood as the violation of a norm. It seems that if one consciously discounts evidential considerations, one simply opts out of the 'game' of belief-formation. These considerations seem to me to show that while it makes sense to speak of truth as the (maybe constitutive) *aim* of belief-formation, invoking truth as a *prescriptive norm* that governs belief is rather pointless. Someone who seeks to make up her mind about whether she should believe that p intends to find out *whether* p is true; the putative norm that she should believe that p iff p is not going to be of any use to her.

misinterpret our evidential situation, then it always possible for us to judge improperly.

How are we to understand the idea that we are afflicted with a pervasive propensity to make empirical judgments on the basis of a mistaken conception of what our theoretical reasons or evidential grounds are? As a first approximation, we can take into account Kant's appeal (CPR, A 52/B 77) to the "empirical conditions under which our understanding is exercised" as including "the influence of the senses, the play of imagination, the laws of memory, the force of habit, inclination, etc., and so all sources of prejudice from which this or that cognition may arise or seem to arise". But an appeal to the factors that Kant enumerates here may seem too thin or abstract, maybe even too trivial. Sure, it is always conceivable that we may misremember things, that we may be in the grip of prejudices, or that our judgment may be corrupted by concealed wishful thinking. But these and similar factors hardly account for a serious possibility of error that *characteristically* afflicts our thought. It seems that there are many occasions for beliefs about mundane matters of fact where the possibility that we may misunderstand our evidence seems purely academic. In response to this worry, I shall mention three issues, each of them crucial to Kant's theory of knowledge, which motivate the claim that our doxastic agency is pervasively and characteristically afflicted with the threat of error.

First, Kant's epistemology is radically *holistic*. It assumes that individual perceptions depend for their evidential significance on their place within an extensive framework of further perceptions, concepts, and laws.⁴⁶⁹ This assumption is based on the idea that the disparate parts of nature that our various perceptions disclose to us are systematically connected. Keeping in mind the holistic character of Kant's epistemology is important for understanding two points that have emerged in the previous section. It explains both why Kant holds that there is a reflective gap between receiving perceptual input and judgment and why Kant endorses a norm that demands that our judgments be based on active, explicit reflection. If propositions had their confirmation conditions separately and in isolation from one another, the need for active reflection would be greatly diminished: having sufficient evidence for accepting a certain proposition would reduce to the reception of those perceptual data that serve as its confirmation conditions. Correspondingly, on such an 'atomistic' model it would be hard to see why a thinker must take active epistemic care to form an adequate conception of her evidential situation: having a perception would leave the subject with no reflective option and *cause* her to believe things about the world.⁴⁷⁰ By contrast, Kant's holism requires that a thinker actively works out what, if any, implications a given perception has in the light of her overall system of cognition. In this spirit, Kant says that "we must endeavor, wherever possible, to bring systematic unity into our knowledge" (CPR, A 650/B 678). This holistic pursuit of empirical knowledge is a complex matter that requires an informed grasp of a picture of nature as a whole. Given the complexity of

⁴⁶⁹ For a very suggestive summary version of Kant's holism, consider what Friedman says in his discussion of Carnap (1999, p. 99): "...it is of the utmost importance that Carnap's conception of knowledge and meaning is Kantian...in that it is 'holistic' rather than 'atomistic'. Concepts do not derive their meaning 'from below' – from the ostensive contact with the given. Indeed, such merely ostensive contact with the given is the very antithesis of truly objective meaning and knowledge; for objective meaning and knowledge can only be derived 'from above' – from the formal or structural relations within the entire system of knowledge."

⁴⁷⁰ The atomistic picture must be assumed by those philosophers of perception who think that in an ideal case of empirical belief formation, the subject of thought is *caused and compelled* to have a belief by external objects that provide her with perceptual input: "A person's direct perception of the way things are in the world around him provides a paradigm case of the integration of causation and rationalization constitutive of compulsion in thought by reason" (Brewer 1995, p. 17).

this task, we must take great care not to misconstrue our evidence, and there is plenty of room for epistemic prejudices and faults to corrupt our thought.⁴⁷¹

The second point worth noticing here is the pervasiveness of empirical illusion. For Kant, empirical illusion is not just a feature of certain special cases, for instance, when we see a straight stick as bent in the water. Rather, Kant holds that our perceptions *always* present empirical objects in ways that depend on the contingent, subjective constitution of the perceiver ("the special character of the subject", CPR, A 820/B 848; cf. section 2 above). When we are affected by outer objects, we cannot help but having "sensations of colors, sounds, and heat do not of themselves yield knowledge of any object. They cannot rightly be regarded as properties of things, but only as changes in the subject, changes which may, indeed, be different for different men" (CPR, A 29/B 45). Even an objective property such as motion is displayed by our senses in a distorted fashion: "The course of the planets is represented to us by the senses as now progressive, now retrogressive" (Prol, 4: 290-1). So, on Kant's view our senses constantly give us subjective impressions that they present as being on par with representations of objective properties such as shape and size. This shows that we are globally susceptible to empirical illusion, *whenever* we represent a particular object. Moreover, this susceptibility is *persistent*; it continues to affect us even once we see through sensory illusions. If we recognize, in hindsight, what led us to make some empirical judgment,

without having to take account of the character of the object, we expose the illusion and are no longer deceived by it, although always still in some degree liable to come under its influence, in so far as the subjective cause of the illusion is inherent in our nature. (CPR, A 821/B 850)

Even once we realize that colors and sounds are mere subjective appearances, we must constantly bring this realization to the foreground of our reflective consciousness. This is the only way to resist our tendency to assign objective evidential weight to subjective sensory appearances of colors and sounds (etc.). I think that Kant's point here is very plausible: if it is indeed correct that our perception of colors, tastes, sounds and apparent motions do not reflect the real (empirical) constitution of the object, then it will require a lot of care not to 'give in' to

⁴⁷¹ The link between a holistic epistemology and an 'agential' picture of assenting to propositions is explicit in Quine, who moves from his rejection of an 'atomistic' conception of empirical confirmation to the idea that we have the liberty of "choosing one statement rather than another for revision in the event of recalcitrant experience" (1951, pp. 39-40; Quine does not like concepts such as 'belief' or 'proposition'). As we saw, for Kant, there is no such thing as *choice* involved in our freedom of empirical thought. Kant's holism is very different from Quine's: Quine's appeal to the idea that we are free to choose between statements reflects his notion that the standards here are ultimately based on *pragmatic* considerations, for instance, on whether a given researcher is inclined to simplicity rather than conservatism (see Quine 1986, pp. 620-1). By contrast, for Kant the attempt to construct a body of empirical knowledge from given sensory experiences requires that there are non-pragmatic, objective principles of theoretical reason that govern the holistic pursuit of empirical truth. Further features that render Kant's holism unique emerge when we contrast his stance on *revisability* and *analyticity* with Quinean and Carnapian holism. Quinean holism characteristically holds (1) that *every* statement is in principle revisable and (2) that *therefore* there are no analytic statements. (2) rests on the equation of analyticity with unrevisability; rejecting this equation gives us a second form of holism, which we might call Carnapian and which accepts some rough analogue of (1) but rejects (2): as Friedman has shown (1999, p. 33), Carnap accepts that analytic 'framework principles' *can* be revised in the light of scientific discoveries. Kant rejects not only (2) that there are no analytic statements, but also holds that there are certain (analytic or synthetic) principles that are immune to rational revision; Kant does, however (at least at times), seem to accept something along the lines of (1) with respect to empirical statements or concepts (see CPR, A 728/B 756).

those perceptions even for someone who has recognized their deceptive nature.

This brings me to my third and most complicated point. Given the way our sense organs work, we cannot help but perceiving objects as red, sour, etc.; but *that* our sense organs work in this way is a contingent fact about our sensibility. By contrast, there is a kind of illusion that is independent of sensibility, that has its source in the very nature of theoretical reason (i.e., which "reason must necessarily encounter in its progress", CPR A 422/B 450) and (hence) that would afflict all discursively thinking beings: namely, *transcendental illusion*.⁴⁷²

This illusion derives its force from two aspects that are crucial to our finite, discursive thought about empirical reality. First, any natural object or event that we encounter in our cognitive pursuits is *conditioned*: it depends for its existence, character or occurrence on numerically different things that constitute its metaphysical and explanatory conditions. Second, theoretical reason strives for *completeness* in its pursuits: whenever it is confronted with something to be understood, such as a natural phenomenon, reason forms the idea of a complete or absolute totality of conditions (CPR, A 308/B 365; A 462/B 490). This idea is cognitively innocent (even, Kant thinks, indispensable). But it produces

what, though a mere illusion, is none the less irresistible, and the harmful influence of which we can barely succeed in neutralizing even by means of the severest criticism. (CPR, A 642/B 670)

Now, the 'harmful influence' that the idea of nature as an absolute totality exerts on our theoretical thought has many faces. I want to sketch three forms that this influence may take.

The first and most dramatic effect of transcendental illusion occurs when we give in to the strong propensity to *reify* the notion of a complete totality of conditions for a given conditioned, that is: when we posit an entity that exists independently of our cognitive effort, a complete whole of reality that contains all the conditions for a given observed object or event (CPR, A 487/B 515). Here it is important to note that *how* one reifies the idea of a complete totality of conditions depends on whether one happens to be more attracted to empiricist or rationalist commitments and principles. The empiricist accepts that natural phenomena can be explained and conditioned only by further natural phenomena. She is thus led to infer from the cognition of a natural phenomenon to the existence of an *infinite* series of empirical conditions ("the chain of the natural order", CPR, A 470/B 498) which cannot terminate in a part that is not itself a conditioned natural phenomenon (CPR, A 469-70/B 497-8). This empiricist conception of

⁴⁷² For the distinction between empirical and transcendental illusion, see CPR, A 296/B 352. For the necessity and the indestructibility of transcendental illusion, see CPR, A 421-2/B 449-50. I think that it is primarily the necessary, *non-arbitrary* (CPR, A 462/B 490) character and the source of transcendental illusion that sets it apart from empirical illusion – that is, the fact that the former kind of illusion is, unlike the latter, rooted in our rational faculties as such independently of any contingent empirical conditions. (Correspondingly, the influence of transcendental illusions is to be countered by no less than a critique of metaphysics.) Apart from these points, there are significant parallels between the two kinds of illusion: Kant holds (CPR, A 298/B 354-5) that *both* transcendental *and* empirical illusion are such that the *Schein* they create does *not* go away even if one sees through them; and in both cases one can (if one takes sufficient care) prevent that they deceive, even though one shall remain always to some extent liable to such deception. (By contrast, the *logical illusion* exhibited by fallacious inferences, which result from a careless failure to attend to logical rules, vanishes completely once it is exposed as illusion; see CPR, A 296-7/B 353.) These parallels explain why Kant uses (CPR, A 644-5/B 672-3) the example of an optic (mirror) illusion to illustrate the nature of transcendental illusion. As Grier emphasizes in her pioneering 2001 discussion, Kant holds that the transcendental illusion itself (freed from its deceptive character) is indispensable for the pursuit of empirical truth.

an absolute whole of conditions for a given conditioned leaves no room for the possibility of there being any uncaused causes or first beginnings. The rationalist, by contrast, argues that this picture does not make sense unless we can conceive of the series of conditions as being grounded in some privileged member: on this basis, she will posit, as part of the world of experience, an uncaused cause or a first beginning that brings closure to the chain of causes or temporal states.

Despite the enormous differences in their respective outlook, the naturalist and the rationalist are led to these outlooks by the same fallacious type of inference: they both infer from something that is given in experience – a particular event or object – to something that is *not likewise* given in experience: namely, a complete series of conditions, conceived either as an infinite or as a finite totality. This inference is epistemically unjustified in that it fails to accord with our objective evidence: we cannot know whether the complete totality of conditions for a given natural object or event is finite or infinite, because this totality is not an object of our experience (CPR, A 482-4/B 510-2). Moreover, this inference has disastrous effects on the use of our reason on the whole: the rationalist picture, which posits theoretically elusive uncaused causes or first beginnings as explanatory conditions, violates the unity of experience and destroys our understanding of nature; the empiricist picture, which declares that *all there is* must be part of the natural order, undercuts our practical self-conception since it deems free, responsible agency and God impossible (CPR, A 472/B 500).

The second effect that transcendental illusion has on the empiricist's and the rationalist's pursuit of empirical truth comes to light in their stance toward the apparent purposiveness or teleological structure of natural organisms. The rationalist will be inclined to hold that nature is *intrinsically and objectively* teleologically structured in a way that derives from the intentions of an all-wise being and that rivals explanations of objects and events according to mechanistic laws (CPR, A 689-702/B 717-30). The empiricist will be inclined to give strict and absolute priority to mechanistic patterns of explanation, and hence she will deny that there is a legitimate place for teleology in our thinking about natural organisms; a denial which, Kant claims (CPR, A 651/B 679; KU, 5: 183), hinders our understanding of nature insofar as it provides an obstacle to the achievement of systematic unity in our empirical cognition.

There is a third negative effect that transcendental illusion has on empirical enquiry. The general idea of a complete totality of nature, or the corresponding idea of the widest possible expansion of our theoretical cognition (CPR, A 643/B 671), comprises a number of different, more specific ideas (CPR, A 652/B 680 ff.). For instance, it contains the idea that the various different events and substances we behold in nature can be systematically understood by reference to *common* properties (the systematic unity of all empirical concepts; CPR, A 652/B 680); but it also contains the idea that the things that we group together under common concepts diversify into different species, i.e., that there are ever ongoing differences among things we classify as having properties in common. Now, the crucial point for the purposes of our discussion is that it may be hard to combine these two ideas *if* one believes (under the influence of transcendental illusion) that they could only relate to the inner constitution of nature, as it is in itself. That is, it may be hard to combine the notion that nature is *intrinsically* diverse (so that every unifying ground masks a potential for further specification) with the notion that nature is *intrinsically* unified (so that all diversity and specificity can be traced to common laws or properties).⁴⁷³ If one is 'speculatively', that is, rationalistically inclined and believes in the

⁴⁷³ Kant holds that considered as objective principles, the ideas of unity and of specificity "could be contradictory"

ultimate unity of nature (CPR, A 655/B 683), one may feel committed to explaining away as merely apparent a putative diversity among natural phenomena; if one is empirically inclined and believes in the ultimate diversity of nature, one may feel committed to dismissing, as merely apparent, a putative unity among diverse phenomena. For instance, someone committed to the idea of intrinsic manifoldness of nature may be inclined to dismiss the search for shared features between a newly discovered species and known species as based on prejudice; conversely, someone committed to the idea of intrinsic unity of nature may see no point in engaging in the search for further diversifications of known causal laws and powers. Both researchers will be hindered in their pursuit of empirical truth:

When we observe intelligent people disputing in regard to the characteristic properties of men, animals, or plants -- even of bodies in the mineral realm -- some assuming, for instance, that there are certain special hereditary characteristics in each nation, certain well-defined inherited differences in families, races, etc., whereas others are bent upon maintaining that in all such cases nature has made precisely the same provision for all, and that it is solely to external accidental conditions that the differences are due, we have only to consider what sort of an object it is about which they are making these assertions, to realize that it lies too deeply hidden to allow of their speaking from insight into its nature. The dispute is due simply to the twofold interest of reason, the one party setting its heart upon, or at least adopting, the one interest, and the other party the other. The differences between the maxims of manifoldness and of unity in nature thus easily allow of reconciliation. So long, however, as the maxims are taken as yielding objective insight, and until a way has been discovered of adjusting their conflicting claims, and of satisfying reason in that regard, they will not only give rise to disputes but will be a positive hindrance, and cause long delays in the discovery of truth. (CPR, A 667-8/B 696-7)

Here, the influence of transcendental illusion manifests itself in the following assumption: if the ideas that we rely on in our theorizing about nature (ideas such as the systematic unity of diverse phenomena or the diversity of phenomena that fall under the same general concepts) have any legitimacy at all, they must relate to the objective inner constitution of nature. It is worth emphasizing that it is the holistic character of empirical knowledge that renders this assumption especially harmful: because the evidential significance of a given perception rests on its place within a wide web of perceptions and concepts, cognitive responses to given perceptions will be negatively affected by the subject's unpreparedness to take into account the relations of unity between the various members of this web; on the other hand, an exclusive focus on relations of unity that comes at the expense of anticipations of diversity will be a stumbling block to the expansion of the web and thereby to the growth of our understanding of the natural order.

To clarify: in leaving out of account the diversity or the unity of the different parts of nature as a constraint on doxastic deliberation and on empirical judgment formation, the subject of thought does not commit a *factual* mistake – her fault is not that she fails to recognize an objective feature of empirical reality. Rather, she violates a *normative* demand of theoretical

(CPR, A 666-7/B 694-5). One may wonder what exactly the contradiction is supposed to be; Guyer 1990, p. 32 argues that the conflict could at most be practical, relating to limits on time and resources. I suggest that what Kant might have in mind – and what might be implied by his rather careful phrase, "could be contradictory" – is the following point: if a thinker proceeds on the assumption that nature is intrinsically and objectively unified or diverse, the *source* of such an assumption is likely to be a more general, rationalistic or empiricist *worldview* (as exhibited, respectively, by the theses or antitheses of the Antinomies) according to which we can only have *either* ultimate unity *or* ultimate diversity in nature because what grounds or explains the one rules out the other (or, at least, according to which it must be held that if either unity or diversity is accepted as constitutive of nature, one has no reason to believe that the other obtains in nature as well).

reason, namely, a norm that "prescribes that we ought to study nature *as if* systematic and purposive unity, combined with the greatest possible manifoldness, were everywhere to be met with, *in infinitum*" (CPR, A 701/B 729). Kant holds that the ideas of theoretical reason, such as that of a complete totality of nature, of a teleological structure of nature, or of unity and manifoldness of natural phenomena, are perfectly legitimate if they are understood as regulative ideas that have a normative point in that they prescribe how we ought to form our judgments about the natural world. We make an illegitimate use of these ideas if we come under the influence of transcendental illusion and 'hypostasize' them as being constitutive of mind-independent reality, that is, describing the inner character of nature (CPR, A 643-5/B 671-3).

A lot more could be said here about Kant's conception of the nature and effects of transcendental illusion. But the discussion I have provided shows why Kant thinks that due to the influence of transcendental illusion, we are prone to think about reality in a subtly but gravely distorted fashion. This tendency is rooted in the nature of theoretical reason, and this provides a transcendental ground for the possibility of theoretical error.

To sum up the discussion of this section: the demanding, holistic character of empirical knowledge, the global presence of empirical illusion, and the ineradicable force of transcendental illusion jointly motivate the idea that finite thinkers are afflicted with a pervasive propensity to misinterpret their theoretical reasons. According to Kant, constant active reflection is the only way in which we can guard ourselves against carelessly giving in to this propensity.

5: Freedom of belief and freedom of choice

We are now in a position to compare and contrast Kant's notions of freedom of will and freedom of thought. My argument so far shows that at an abstract level, the two kinds of freedom can be characterized in just the same way. Both involve, *negatively speaking*, the absence of determination by sensible causes. Moreover, both involve the *positive* capacity to form our attitudes on the basis of our awareness of objective reasons; and in both cases this positive ability has a shadow, namely, our pervasive propensity to *fail* to respond to objective reasons.

The similarities between theoretical and practical freedom come out nicely if we compare what Kant says about the role that sensibility plays for the *misuse* of our rational capacities:

If an appearance is given us, we are still completely free as to how we want to judge things from it. (...) And then it is not the fault of the [sensible] appearances at all, if our cognition takes illusion for truth... (Prol, 4: 290)...the source of the bad cannot lie in any object that determines the Willkür through inclination, or in any natural impulse, but only in a rule that Willkür will makes for itself for the use of its freedom, that is, in a maxim. (RGV, 6: 21)...the source of this badness cannot, as is usually done, be placed in the *sensibility* of man and the natural inclinations springing therefrom...[it is rather] found in him as a freely acting being. (RGV, 6: 38)

Although sensibility may influence us to think and choose incorrectly, it is not a *source* of either practical or theoretical failure. In the theoretical case, the senses do not err: an improper theoretical judgment may occur under the influence of sensibility (i.e., the influence of perceptual input or the play of the empirical imagination) but is to be imputed to the freely thinking self who allows herself to be misled, namely, to the free exercise of her faculty of understanding. In the practical case, the senses are not evil or bad: a bad choice is the result of the free exercise of a person's faculty of volition, which assigns undue weight to her desires.

However, the fact that practical and theoretical freedom for Kant are structurally identical

in certain general respects should not lead us to overlook that these two forms of freedom also differ specifically from one another in important respects. I want to draw attention to two such differences. The first one concerns, again, the role of sensibility. Our theoretical faculties *need* the contribution of sensibility for their proper exercise, because there is no empirical cognition without the data that we receive through our senses. Sensibility is both a (potential) defeater and a (necessary) enabler of empirical judgment: the influence of the senses on empirical judgment may result – if we are careless – in empirical error, but it is also a necessary condition for the possibility of the proper exercise of our cognitive capacities that results in empirical cognition. By contrast, the proper exercise of our free practical faculties does *not* (with regard to *moral* agency) depend on the contribution of sensibility: here sensibility plays solely the role of a potential defeater. There is nothing in moral thought or agency that essentially depends on a contribution of receptivity or sensibility. As far as morally good choices are concerned, the influence of sensible inclinations is a mere obstacle that pure practical reason must remove:

The essential point in every determination of the will by the moral law is that being a free will it is determined simply by the moral law, not only without the co-operation of sensible impulses, but even to the rejection of all such. (KpV, 5: 75)

Now, for Kant, the exercise of our practical capacities relates not only to the formation of choice but also to practical reasoning that issues in the formation of practical judgments. And with respect to those practical judgments that Kant considers an expression of our freedom, namely, moral judgments, we can say that practical reason legislates its (moral) laws in absolute spontaneity, without being constrained by our sensible, receptive nature:

[Practical] Reason does not here follow the order of things as they present themselves in appearance, but frames for itself with absolute spontaneity an order of its own...to which it adapts the empirical conditions, and according to which it declares actions to be necessary although they have never taken place, and perhaps never will take place. (CPR, A 548/B 576)

Thus, one central difference between practical and theoretical freedom is that the free exercise of our epistemic faculties does, whereas the free exercise of our practical faculties does not, depend on the deliverances of sensibility.⁴⁷⁴

The second difference between practical and epistemic agency can be brought out by

⁴⁷⁴ Here I disagree with Irwine 2007 and Longuenesse 2005, who both claim that moral judgments for Kant depend, like empirical judgments, on the interplay of receptivity and spontaneity. Irwine's claim is based on the rather confused attempt to read Kant as a contemporary moral realist. For Longuenesse, the point seems to be that what our moral judgments deem permissible or impermissible are the choices which incorporate the various desires that affect our receptive or sensible nature into general volitional rules (or maxims). I agree that this idea is central to Kant. But here our empirical desires do not play the role for practical judgment formation that sensible intuitions play in empirical judgment formation. An empirical judgment is (partly, but essentially) *based on* an empirical intuition, where this means that the intuition yields part of the judgment's evidential ground and truth-condition. Moral judgments are not likewise based on empirical desires, or rather: if they are, this counts against their characteristic validity. A judgment that was (surreptitiously) influenced by an empirical inclination would lose its claim to expressing a *necessary* demand of reason or to being a categorical imperative (GMS, 4: 419). A related point is that a judgment based on sensible intuition is valid only for a being that has the relevant form of sensibility or receptivity, whereas moral judgments are, as Kant never tires of emphasizing, valid for all rational creatures including those whose forms of receptivity differs from ours or that (like God) have no forms of receptivity at all. The possibility of practical legislation is not essentially tied to the possession of a sensible or receptive faculty. (See my chapter VIII, section 5 for a more detailed discussion of this point.)

considering the manner in which our rational imperfection manifests itself in each case. As we saw, we have a pervasive tendency to choose contrary to our reflective awareness of what our practical reasons are, but we cannot believe contrary to our reflective sense of what our theoretical reasons are. This is one side of the story. Now, Kant thinks that our normative practical self-awareness typically is as it ought to be; at least the fundamental principles of morality leave an inescapable mark on our deliberative consciousness, which prevents a lasting corruption of our awareness of what we morally ought to do⁴⁷⁵:

It is only necessary to analyze the judgment that men pass on the lawfulness of their actions, in order to find that, whatever inclination may say to the contrary, reason, incorruptible and self-constrained, always confronts the maxim of the will in any action with the pure will, that is, with itself, considering itself as a priori practical. (KpV, 5: 32) ...the voice of reason in reference to the will [is] so clear, so irrepressible, so distinctly audible, even to the commonest men... (KpV, 5: 35)

By contrast, the voice of theoretical reason is not nearly as audible or incorruptible: there is no persistent call of epistemic consciousness. This is shown by the fact that we typically judge the natural world without reflecting and end up believing that wine is red and sweet or that reality is a finite or an infinite totality of conditions. In the theoretical case, our rational imperfection *secretly* gets a hold on us: it occurs in the form of an unobserved corruption of the normative self-awareness that underlies our cognitive efforts (CPR, A 294-5/B 350-1).

Hence, there is a double asymmetry between practical and epistemic agency: on the one hand, our reflective awareness of how we ought to choose is characteristically well-informed whereas our reflective awareness of how we ought to judge nature is characteristically misinformed. On the other hand, our choices characteristically fail to accord with our reflective sense of how we ought to choose, whereas our judgments about nature characteristically accord (for better or worse) with our reflective sense of how we ought to judge nature.

6: Freedom of thought and the causality of nature

As we saw, many commentators deny that Kant adopts a conception of epistemic freedom or spontaneity that is even remotely analogous to his conception of practical freedom or spontaneity. In this chapter I have shown that this kind of interpretation rests on a mistake, especially insofar as it is based on a fallacious inference from the insight that Kant rejects doxastic voluntarism to the notion that Kant has no room at all for a notion of freedom of thought or epistemic agency. We also saw that many commentators hold that Kant's conception of epistemic spontaneity is compatible with causal necessitation and may thus be naturalized. I want to conclude by at least indicating that Kant's conception of the freedom involved in the formation of empirical belief is as anti-naturalistic and incompatibilist as his conception of the freedom involved in the formation of choice: the 'I that thinks' is not a thinking mechanism.

There are two ways of making this point. The first one is the straightforward upshot of the argument of this chapter. I have shown that for Kant, the possibility of error is a pervasive feature of our theoretical thought, but cognitive errors are never quite forced upon us. In our cognitive efforts, we must consider ourselves imperfectly rational thinkers: on the one hand, our rational powers always enable us to correctly interpret our evidential situation; but on the other hand, the imperfections inherent in our rational powers always leave us with the option to

⁴⁷⁵ This does not mean that an unobserved corruption of moral judgment via the influence of sensibility cannot happen: desire and prejudice may lead us to misapply the formulae of the categorical imperative incorrectly.

misconceive our theoretical reasons. By contrast, a thinking mechanism is always causally determined to respond to external stimuli in one particular manner: it is necessitated either to function or to malfunction. Correspondingly, our thought, unlike a mechanistic process, stands under normative laws that we may comply or fail to comply with, and that we represent to ourselves in the form of 'ought'. Here, no less than in the practical case (cf. chapters IV and VI), the point of the formula 'ought' is to represent the right course of action (in a broad sense of 'action') to a being that may or may not take this course. We have encountered a number of such epistemic 'oughts' in the course of this paper: for instance, the demand that we must not form our doxastic attitudes without reflecting; or the demand "that we ought to study nature *as if* systematic and purposive unity...were everywhere to be met with".

This conception of our epistemic agency is hard to square with the idea that we are always necessitated to judge in one particular way. Compatibilists offer interpretations of 'can' that would allow us to say that a thinker can, or could have, judge(d) otherwise even if she was causally necessitated to make a particular judgment. Kant rejects such compatibilist analyses of what is possible; and as I have argued elsewhere (cf. chapter VI, section 2.1), it is hard to see how such analyses could convince us that, for instance, someone who forms the faulty belief that the stick in the water is bent has the genuine option to avoid this belief if her thinking is exclusively governed by deterministic neurological laws. So, one reason why we should reject the mechanistic interpretation is that according to Kant, we as imperfectly rational thinkers have, whereas a mere mechanism lacks, alternative possibilities; correspondingly, our theoretical thinking is, whereas the processes of a mere mechanism are not, governed by epistemic 'oughts'.

This route toward anti-naturalism or incompatibilism about freedom of thought derives from our self-conception as imperfectly rational thinkers. I want to mention a second route toward Kant's incompatibilist about freedom of thought which which is more important to Kant himself. This point has been implicit in the argument of this chapter. We saw that for Kant the core of our freedom of thought is our capacity to judge and to form our doxastic attitudes in response to objective theoretical norms and reasons. But, Kant thinks, if our judgments were exclusively determined by natural causes, they would not be responsive to objective reasons.

Why should this be? I want to bring out what (I think) Kant's point is by reference to a wholly anti-Kantian conception of the relation between mind and reality. On this anti-Kantian picture, it is hard to see why determinism should pose a threat for our self-image as rational cognizers. After all, our empirical beliefs try to fit the world: so what could be bad about the idea that our beliefs are caused by precisely those empirical objects that they aim at representing in the first place?⁴⁷⁶ To see why Kant would find this picture unacceptable, we can focus on the implications that it has for our world-directed *causal* thought. For Kant, the concept of causality is modally thick: it involves the representation of nomic necessity. Now, Kant agrees with Hume that we do not find anything that answers to this modal feature in the content of our perceptions. So, if we were a natural mechanism, our compulsion to think in causal terms would not come from the external world and the perceptual stimuli it gives us, but from the internal design that is implanted in us as a matter of contingent empirical fact. Our disposition to think in causal terms would exclusively depend on the way in which our minds or brains happen to be wired. And for Kant, this would mean that our thinking in causal terms is wholly arbitrary:

⁴⁷⁶ This line of thought is most explicit in Nagel 1989 and Wiggins 1969, 2003. Their slogan is "determination by the truth". See chapter XII for a detailed discussion of the line of thought embodied in this slogan.

[It] may be proposed that...the categories are subjective dispositions of thought, implanted in us from the first moment of our existence... [But] the concept of cause...which expresses the necessity of an event under a presupposed condition, would be false if it rested only on an arbitrary subjective necessity, implanted in us... I would not then be able to say that the effect is connected with the cause in the object...but only that I am so constituted that I cannot think this representation otherwise than as thus connected. This is exactly what the skeptic most desires. (CPR, B 167-8)

The 'merely subjectively determining causes' Kant mentions here are the kinds of psychological (or, as we might prefer to say, neurological) causes and conditions that regulate the processes of the empirical mind. 'Subjective necessity' is a term that Kant constantly uses when he imagines a scenario in which we are determined to think in certain ways by contingent conditions over which we have no rational control, such as psychological (or, as we might say, neurological) laws. Consider, for instance, his appeal to scenario in which our understanding stands under the "mechanism of merely subjectively determining causes" (RezSchulz, 8: 14; cf. p. 195 above).⁴⁷⁷

Kant's rejection of the idea that the categories have their origin in "dispositions of thought, implanted in us from the first moment of our existence" tells against the functionalist claim that Kant is not adverse to the idea that we might be thinking mechanisms whose judging in causal terms is a matter of us being wired by our internal 'hardware'; and this rules out the kind of naturalistic interpretation of theoretical spontaneity that is advanced, for instance, by Sellars and Kitcher. Sellars rightly points out that Kant's conception of practical spontaneity is anti-naturalistic because practical reason is a capacity to judge and choose in ways that are independent of our contingent empirical psychology.⁴⁷⁸ But he does not recognize that for Kant, the understanding is completely on par with practical reason in *this* respect: the understanding is also a capacity to judge in ways that are independent of our contingent empirical psychology.

This is perfectly consistent with the previously mentioned fact that the understanding is limited by the empirical in a way that practical reason is not. We must distinguish here between two points: on the one hand, pure theoretical concepts such as causality can be legitimately *applied* only to the empirical world; on the other hand, these concepts and the judgments that

⁴⁷⁷ Kant here standardly has in mind Hume's idea that our thinking in causal terms has its origin in the contingent 'habits' of the phenomenal, empirical mind. See CPR, B 5, B 167-8, A 94/B 126-8; and KpV, 5: 12-13.

⁴⁷⁸ Sellars 1970 holds that the spontaneity required for morality is absolute whereas the spontaneity required for empirical thought is only 'relative' because practical reason, unlike theoretical reason, has an 'intrinsic premise' that is not given to it "from without". What he means is that practical reasoning for Kant cannot be understood in terms of the Humean 'desire in, desire out' principle. But he overlooks that for Kant theoretical reasoning cannot be understood as a mere function from contingent psychological input either. Note, moreover, that Kant himself never uses the notion of a merely relative spontaneity. This notion is an invention of Sellars that has found a steady place in the secondary literature. Following Sellars, Allison 1990 holds that unless we had free will, we would have mere relative spontaneity: we would be "an automaton spirituale...a thinking mechanism" (*ibid.*, p. 63; compare p. 228). Kant himself does, derisively, use the notion of automaton spirituale in discussing the Leibnizean conception of 'comparative' freedom of will which holds that our actions are the result of psychological causes that are themselves causally determined. But Kant *never* applies the notion of an automaton spirituale to the spontaneous understanding. He wants to *rule out* that our synthesizing and thinking in terms of the concept of causality is the upshot of psychological laws (Hume's customs; see the preceding footnote). Moreover, when Kant entertains the possibility that our minds might be mere thinking mechanisms (KpV, 5: 101), he says that our self-consciousness as spontaneous beings would be tropistic, a 'mere illusion'; and this is of a piece with his claim if we were so constituted that we could not help but thinking in causal terms, the concept of cause would be invalid (CPR, B 167-8). So the spontaneity of the understanding that is responsible for the representation of causality is *incompatible* with the idea that we are thinking mechanisms.

employ them *do not have their origin* in our empirical psychology:

The empirical is only the condition of the application...of the pure intellectual faculty. (CPR, B 423) If...a judgment...asserts a claim to necessity, then...it would be absurd to justify it by explaining the origin of the judgment psychologically. For...if the attempted explanation were completely successful it would prove that the judgment could make absolutely no claim to necessity, precisely because its empirical origin can be demonstrated. (EEUK, 20: 238)

So, our *application* of the concept of causal necessity is limited to the empirical world; hence, our causal judgments depend on the reception of sensible data that disclose the empirical world to us. But our causal thoughts do not have their origin in, and are not produced by, empirical, psychological causes such as Hume's principles of custom or habit.⁴⁷⁹

Hence, for Kant the pure, empirically unconditioned understanding is the author of theoretical laws, such as the principle of sufficient reason; just like pure, empirically unconditioned practical reason is the author of the categorical imperative. Accordingly, Kant deems both the understanding and practical reason autonomous; he refers to "...[the autonomy] of the understanding, in view of theoretical laws... or [the autonomy] of reason, in practical laws..." (EEUK, 20: 225). Surely, a causally determined thinking mechanism *cannot* be considered autonomous.⁴⁸⁰

⁴⁷⁹ One might try to deflate the EEUK, 20: 238 passage by proposing that it can be read as advocating no more than the idea that we cannot justify a priori judgments by appeal to their psychological origins. Kant accepts this idea; but, as the second sentence shows, Kant's anti-psychologism is much stronger than that: he holds, in addition, that if we *could* demonstrate that a judgment had its origin in our psychological constitution, this would impugn the judgment's claim to validity. The second sentence is clearly intended to justify the first one. I first argued this point, in the context of responding to the charge that Kant commits 'the genetic fallacy', in chapter V, section 1.

⁴⁸⁰ Sellars might respond that his conviction that the 'I that thinks' may be a mere mechanism draws on the fact that Kant himself conjectures that the (noumenal) thinking self might be a mere 'thing' which as such (MM, 6: 223) cannot be considered imputable. But here we need to pay attention to the context in which Kant raises the possibility that the thinking self is a mere 'thing': he does so in the Paralogisms, where he seeks to show that we cannot verify (or falsify!) the idea that the thinking self is an imputable person that remains the same through changes on the basis of analyzing the concept of apperception (of the 'I that thinks'). Now, as is well-known, Kant thinks that we have practical grounds that entitle us to a notion of moral personhood that suffices for ascriptions of responsibility: namely, our awareness of our capacity to comply with objective laws of volition. In the same way, he legitimizes the imputation of theoretical judgments by appeal to our self-conception as thinkers who can comply with objective laws of theoretical thought (see again RezSchulz, 8: 14) and whose mistakes are due to their own carelessness (Prol, 4: 290). This constitutes no breach of his criticism of the rationalist pretensions to know the soul as it is in itself on the basis of intellectual intuition or conceptual analysis (see my chapter X, section 7 for further discussion). However, I think that epistemic and moral imputability are different concepts. In particular, the notion of personal identity over time is relevant exclusively for the purposes of assigning practical imputability: for here 'imputability' is essentially linked to praise or blame, and hence it essentially matters whether we have reason to believe that the source of free agency (the noumenal person) underlying an appearance we are addressing now is identical to that which was underlying the appearance who performed a past act that we consider praise – or blameworthy. By contrast, in the epistemic case, praise and blame – and hence the idea of numerical identity over time – play no substantive role. In fact, the *only* point of the ideas of imputability and spontaneity in the epistemic sphere is to ensure that it is always within the power of a free, spontaneous subject of thought to form her beliefs on the basis of objective theoretical reasons. No thick notion of personhood is implied or required here. This explains why Kant in the *Metaphysics of Morals* says that imputability requires free will: he is concerned with the kind of imputability exhibited in the formation of free choices, attitudes which display (unlike the free doxastic states connected to the activity of free judging) respect or disrespect for persons, and which merit praise or blame for this reason alone.

Chapter X: Synthesis, Spontaneity, and Normativity

In the preceding chapter, I defended the claim that Kant has a substantive notion of freedom and agency that applies to the formation of empirical judgments and doxastic attitudes (such as 'assent'). In this chapter, I will be concerned with a different dimension of Kant's account of epistemic activity: namely, with the kind of synthesis of intuitions that epistemically precedes the formation of conceptually determinate judgments about nature. I examine whether, and to what extent, this form of epistemic activity can be considered free and norm-governed.

1: The original synthesis of intuitions: descriptive and normative approaches

I want to begin by summarizing, very briefly and inconclusively, how Kant conceives of the kind of synthesis that is epistemically prior to the synthesis of empirical concepts in a fully determinate judgment of experience (e.g., 'some trees bloom in winter').

In the Transcendental Deduction (CPR, B 130-1), Kant states that 'analysis' presupposes 'synthesis'. Here, 'analysis' is not limited to analytic judgments, but quite generally relates to the kind of reflection that forms clear and distinct concepts out of given representations.⁴⁸¹ Before a discursive understanding can analyze intuitive representations into concepts, these intuitions must be subjected to an act of combination that renders them suitable elements for such analysis. This act brings given intuitions to *synthetic unity* with one another.⁴⁸² We must attribute this combination to a spontaneous act of the understanding because intuitions are received in an uncombined manner (CPR, B 130-1). The point of such combination is to relate sensible intuitions to objects. Relating intuitions to objects involves representing something distinct from our representations⁴⁸³, something these representations are representations *of* (an intentional object of thought); it is in virtue of being related to objects that sensible intuitions may be referred to something intersubjectively accessible beyond the course of a person's private psychology (and thus satisfy a cognitive precondition for being analyzed into empirical concepts shared by different thinkers). A given representation can acquire such objective significance only if it can be combined (brought to synthetic unity) with other given representations. A representation that could not be so combined, and could not be related to objects, would have no cognitive significance: it would 'at least be nothing to me' (CPR, B 132).⁴⁸⁴ Now, Kant thinks that the synthesis of sensible intuitions that brings them to synthetic unity achieves the goal of relating them to objects because it is directed by the categories. The point of Kant's appeal to

⁴⁸¹ Cf. CPR, B 103/A 77: "Before we can analyse our representations, the representations must themselves be given, and therefore as regards content no concepts can first arise by way of analysis. Synthesis of a manifold...is what first gives rise to knowledge. This knowledge may...at first be crude and confused, and therefore in need of analysis." See Longuenesse 2001, p. 67 for a good commentary on this issue.

⁴⁸² Hence Kant's point that the synthetic unity of apperception is prior to all determinate thought (CPR, B 135).

⁴⁸³ Both versions of the Deduction claim that there is an explicit connection between representing an object and bringing the manifold of intuitions to synthetic unity; compare CPR, B 137 with CPR, A 105: "...since we have to deal only with the manifold of our representations, and since that *x* (the object) which corresponds to them is nothing to us -- being, as it is, something that has to be distinct from all our representations -- the unity which the object makes necessary can be nothing else than the formal unity of consciousness in the synthesis of the manifold of representations."

⁴⁸⁴ A representation that could not possibly be combined with other representations of mine might still 'belong to me' in some psychological sense that lacks cognitive significance. Thus Kant can (pace Guyer 1980, pp. 209-10 and Henrich 1969) accept that the merely psychological (self-) ascription of representations is possible even if those representations are unconscious. See Allison 2004, p. 164.

judgment in § 19 of the B Deduction is that judgment is the fundamental goal toward which the entire use of the understanding is directed. Kant has argued earlier (§ 9) that there is a fixed set of forms or configurations of judgment, and (§ 10) that there is a set of concepts, the categories, that correspond to these forms. If intuitive representations must be subjected to an act of synthesis that renders them suitable candidates for concept and judgment formation, such synthetic combination needs to be informed by rules of synthesis that correspond to the logical structure of judgment. Since the categories correspond to the forms exhibiting this logical structure, they must provide the rules that govern the original synthesis of intuitions.⁴⁸⁵

I take this to be an uncontroversial and uninformative picture of the general point of Kant's appeal to the need for a rule-governed synthesis of sensible intuitions. Among the many questions that this picture raises, the ones I will address in this chapter are: how should we understand the notion of a rule-governed synthesis of sensible intuitions? How does Kant conceive of these rules and of the activity that is governed by them?

A number of recent interpretations interpret Kant as putting forward these rules of synthesis in a descriptive rather than a normative spirit. This contrast requires clarification. For once, we need to distinguish between two different types of descriptive interpretations. One is what I shall call the *naturalistic approach*, which holds that Kant puts forward the rules of synthesis in the spirit of a psychological investigator who tries to explain and understand how the mind works.⁴⁸⁶ The most extreme but also the most sophisticated reading of this kind is offered by Patricia Kitcher, who holds that these rules 'govern' synthesis just as empirical, psychological laws constrain sub-personal processes:

Rules govern synthesis only as the law of gravity governs the movements of the planets: Theorists can appeal to these rules to describe what is happening. (...) Kant doubts that we [as individuals] are aware of any such thing, even though we [as theorists] must assume that it takes place. (1993, p. 83)

A different type of descriptive interpretation has been proposed by Beatrice Longuenesse. She is not adverse to the idea that Kant, in the *Analytic*, is doing psychology⁴⁸⁷, but unlike the proponents of the naturalistic approach Longuenesse does not think that Kant is trying to give us knowledge about actual (empirical) processes. Rather, "what Kant is describing are universal modes of ordering our representations, whatever the empirically determined processes by way of which those orderings occur."⁴⁸⁸ He aims at "the analysis of a specific mode of cognition".⁴⁸⁹

Now, the contrast between such descriptive accounts and a distinctively *normative*

⁴⁸⁵ See CPR, B 143: "All the manifold, therefore, so far as it is given in a single empirical intuition, is determined in respect of one of the logical functions of judgment, and is thereby brought into one consciousness. Now the categories are just these functions of judgment, in so far as they are employed in determination of the manifold of a given intuition."

⁴⁸⁶ Attempts to 'psychologize' Kant's theory of synthesis are not new; in the 19th century, Helmholtz and Lange encouraged such a reading. In the past couple of decades, the psychological reading of Kant's theory of synthesis has taken the shape of a *functionalist* interpretation. Kitcher is the pioneer of this kind of approach; see, among others, her 1984, 1991, 1993 and 1999. Other proponents of the functionalist reading include Brook 1994; Falkenstein 1995; Meerbote 1991; Sellars 1970. Recent writers who explicitly reject the functionalist approach include Allison 1996 and Pippin 1987.

⁴⁸⁷ She suggests (2001, p. 34) that Kant's appeal to synthesis is to be understood, in part, as a "psychological argument (the description of mental acts)".

⁴⁸⁸ Longuenesse 2009, pp. 93-4.

⁴⁸⁹ Longuenesse 2001, p. 34.

interpretation of Kant's enterprise in the *Analytic* is not straightforward. While some proponents of the naturalistic reading altogether reject the notion that Kant's project in the *Analytic* has a normative dimension⁴⁹⁰, both Longuenesse and Kitcher are sensitive to the fact that Kant is concerned with questions about the *validity* of certain concepts (CPR, A 84/B 117). They both hold that Kant's analysis of cognition establishes, via the idea that synthesizing in terms of (e.g.) the concept of causality is a condition of cognition, that we *ought to* synthesize in terms of this concept: thinking in causal terms is vindicated as *correct* by Kant's analysis.⁴⁹¹ Nevertheless, both Kitcher and Longuenesse reject a certain understanding of the idea that Kant's account in the *Analytic* is meant to be normative: they both deny that the justification Kant provides for synthesizing in accordance with the concept of causality is accessible to the subject of synthesis in a way that is relevant to her synthesizing activity. For both Kitcher and Longuenesse, the question about how we ought to synthesize *arises only from the external perspective of the theorist*; subjects synthesize their representations 'blindly'⁴⁹², and the fact that they synthesize under (e.g.) the concept of causality is independent of their conscious awareness that they ought to synthesize in this manner.⁴⁹³ An interpretation that is normative in a sense rejected by Kitcher and Longuenesse claims that Kant seeks to establish prescriptive norms that are meant to stake their claim from within the conscious perspective of synthesizing subjects.

2: Transcendental logic and psychology

We can begin by noting a severe problem that would arise for Kant if his account of rules of synthesis were meant to give us a descriptive account of mental activity. The problem is as simple as it is compelling: Kant affirms both (1) that the spontaneous act of bringing intuitions to synthetic unity is performed by the noumenal mind; and (2) that we do not have theoretical

⁴⁹⁰ Brook notices that "psychological readings of Kant's theory of representation have had a bad press...since they are thought to commit the sin of psychologism ..." (1994, p. 5). He 'avoids' this objection by deeming it 'obviously' misguided: "...when we turn to what the mind must be like to reason and know, we are turning to description, to how the mind does and must work...we are not seeking a normative account of how it should work. In connection with Kant, even that has sometimes been denied, but it seems to me obvious that here we are not doing normative theorizing" (ibid., p. 5). I find it rather obvious that Kant's project is normative in a deeper sense, a reply I shall defend in the course of this chapter (see footnote 500 below for a criticism of Brook).

⁴⁹¹ Kitcher argues that Kant is committed only to what she calls a 'weak psychologism', "the view that psychological facts may be important to... normative claims, even though they cannot establish such claims" (1993, p. 10); in Kant "normative and factual claims commingle" (ibid., p. 21). As I understand her, she means that Kant's theory uncovers the causal processes that are operative in the mind and establishes that the representations produced by them are justified because they are necessary for cognition (cf. her 1999, p. 430).

⁴⁹² For Longuenesse's commitment to this idea, see her 2001, p. 64.

⁴⁹³ Anderson 2001 is a special case: he explicitly conceives of rules of synthesis as normative rules, and I agree with him about many important points (some of which I shall note below). Still, I think that his account is very close in spirit to the accounts offered by Kitcher and Longuenesse. This is because he seems to hold (see especially p. 292) that the sense in which Kant's rules of synthesis are normative is this: we can look at actual representations of any *sort* (even ones that are completely beyond the subject's rational control, such as those induced by a dream) and evaluate them as correct or incorrect by assessing whether "their structure is homomorphic to" a priori rules of synthesis. The model here is (as in Kitcher) that of looking, from a third-person theoretical perspective, at representations as (well- or malformed) products, rather than that of looking, from a first-person 'engaged' perspective, at representations as something to-be-combined in a manner one deems rationally appropriate. The account I favor, and which I shall try to motivate in what follows, seems to me most explicitly endorsed by Paton 1970, who stresses (p. 375 ff.) the significance of the idea that as subjects of synthesis "we must be aware, however dimly, of the character of the rule of the synthesis".

insight into the activity of noumena.⁴⁹⁴ From (1) and (2) it follows that Kant would defeat his own epistemic tenets if he were to offer his account of synthesis as providing us with theoretical insights concerning the noumenal mind, or with a description of noumenal mental acts.

Those who interpret Kant as offering his account of synthesis in the spirit of a psychological theorizer are not usually upfront about this problem and their intended solution, but Kitcher stands out as being admirably clear in her response. While she does not deny that Kant affirms (1) as a matter of textual fact, she argues that Kant is confused here and should accept that spontaneous synthesis is performed by the empirical rather than the noumenal self.⁴⁹⁵ If this revision to Kant's official doctrine is made, nothing stands in the way of reading his account as a theory about how the mind functions. Now, a lot hinges here on whether Kant is in fact incoherent in proclaiming that the synthesizing activity of the thinking self is not cognizable. If so, then a revision of this claim that accords well with other pieces of his doctrine might be the best course to take for a sympathetic reader. But if, as I believe, there is no incoherence⁴⁹⁶, then for the purposes of trying to understand what Kant intends with his appeal to spontaneous synthesis, we should read him under the presumption that he does not aim at giving us psychological insight into how the mind – noumenal or empirical – works.

The best way to assess this presumption must surely be to review Kant's own attitude toward psychology. Kant uses the label 'psychology' in two different senses. First, he acknowledges that there is a scientific study of the mind or soul called 'empirical psychology', which he considers a legitimate discipline, albeit one that suffers from severe methodological difficulties. Second, Kant refers to the non-empirical investigations of the soul carried out by his rationalist predecessors as 'rational psychology'. Kant does *not* officially acknowledge a third sense of 'psychology': it is Strawson, not Kant who invokes the label of 'transcendental psychology'.⁴⁹⁷ Since Kant obviously does not take himself to be engaged in the (as he sees it) imaginary subject of rational psychology, our question must be how he conceives of the relation between his enterprise and the scientific study of the mind by empirical psychology.

Here we need to consider how Kant characterizes his 'logic', i.e., his doctrine of the laws of thought (CPR, A 50-8/B 74-82). He introduces a tripartite distinction between (I) pure general

⁴⁹⁴ (1) and (2) are summarized in Kant's notion that the thinking self (the 'I think', the subject of spontaneous synthesis) cannot become an object of experience. See, for instance, CPR, B 421; B 430. See CPR, B 157-8 for Kant's emphasis that as a subject of transcendental synthesis, I am not aware of myself as I appear to myself' (i.e., as I am phenomenally/empirically), but as 'an intelligence'; this self-awareness relates (GMS, 4: 451-2) to myself as a member of 'the intelligible world', i.e., as a noumenon. See CPR, A 546/B 574 for Kant's emphasis that the subject of pure apperception knows herself as an intelligible object. I discuss these points in section 7.

⁴⁹⁵ Kitcher 1984, pp. 122-26.

⁴⁹⁶ Kitcher's argument for the claim that the subject of transcendental synthesis is the empirical/phenomenal self seems to run as follows: it is a known fact that all representations must be united in one 'I that thinks'; since this is a known fact, it must be known through the categories, and "the doctrine of apperception must present a phenomenal aspect of the self" (1984, p. 123). But here she is begging the question. If Kant were defending the claim, 'all representations must, as a matter of empirical necessity, be united in one I that thinks', then he would be committed to the notion that the doctrine of apperception must concern the empirical self. But *that* Kant is trying to establish such a claim is what Kitcher needs to prove in the first place. If Kant is elucidating a normative condition that may or may not be met in the phenomenal world, then his doctrine of apperception need not concern how things happen to stand with phenomenal self at all. It should also be noted that Kant explicitly says (CPR, B 130-1) that the principle concerning the unity of apperception is *prior to* the categories, and so it is hard to see how he could think that it is known *through* the categories.

⁴⁹⁷ See Strawson 1966. Walker 1978 reminded me of this fact.

logic, (II) applied general logic, and (III) transcendental logic. (I) is 'general' in that it concerns laws of all thought (regardless of whether it involves empirical or a priori concepts). Transcendental logic only concerns laws of thought that involve a priori concepts, or the 'pure' thought of objects (CPR, A 55-6/B 80-1). Kant says that he is not (in the Analytic; contrast his Anthropology) concerned with (II), and he strongly emphasizes the anti-psychological character of (I) and (III): (I), like (III), "has nothing to do with empirical principles, and does not, as has sometimes been supposed, borrow anything from psychology, which therefore has no influence whatever on the canon of the understanding" (CPR, A 54/B 78).⁴⁹⁸

In response to this textual evidence, proponents of the idea that Kant's Analytic tries to discover psychological laws that describe the actual workings of the (empirical) mind typically suggest that it is only a specific kind of (empirical) psychology from which Kant's logic must be dissociated, namely, psychology that is methodologically based on *introspection*. They claim that a kind of psychology that is not limited to this methodological presupposition can inform, or ground, transcendental logic.⁴⁹⁹ But the *interpretive* relevance of this suggestion altogether eludes me. Kant himself clearly does not countenance the possibility of a legitimate empirical discipline of psychology not based on introspection. Thus an appeal to such a discipline can hardly shed light on how Kant conceived of the character of his own enterprise in the Analytic. We should take his declaration of the independence of what he is doing from empirical psychology, in conjunction with his dismissal of rational psychology, to be indicative of the fact that he doesn't take himself to be doing descriptive psychology (aimed at theorizing about the mind) at all.

Maybe Kitcher would respond that the point is rather that Kant *ought* to have countenanced the possibility of non-introspective empirical psychology, and that the most sympathetic reconstruction of his doctrine will show the compatibility of what he is doing and the approach of (say) contemporary cognitive science. Against this, I want to argue that Kant's resistance toward the idea that his transcendental logic draws on (empirical) psychology is not essentially connected to his conception of the (impoverished) methodology of (empirical) psychology. Rather, this resistance is based on the idea that the rules established by transcendental logic and the rules established by (empirical) psychology have an altogether different *point and status*: the latter are empirical (causal) laws, whereas the former are non-empirical laws. To see this, let us consider in more detail the contrast Kant draws between applied logic (II) and general/transcendental logic. He says that (II) is 'impure' because its verdicts depend on the findings of empirical psychology; the purpose of (II), however, is not the same as that of empirical psychology. Applied logic, unlike empirical psychology, serves as a "cathartic of the common understanding" (CPR, A 53/B 78), which is to say that it deals with the question of how the conditions that empirical psychology discovers may further or hinder the proper exercise of the understanding (i.e., its compliance with laws of thought). Kant explains:

What I call applied logic...is a representation of the understanding and of the rules of its necessary employment *in concreto*, that is, under the accidental subjective conditions which may hinder or help its application, and which are all given only empirically. (...) Pure general logic stands to it in the same relation as pure ethics, which contains only the necessary moral laws of a free will as such, stands to the doctrine of the virtues [i.e.] the doctrine which considers these laws under the limitations of the feelings, inclinations, and passions to which men are more or less subject. Such a doctrine can never furnish a true

⁴⁹⁸ It is clear from what Kant says at CPR, A 58/B 82-3 that this point applies to both (I) and (III), i.e., to both general and transcendental logic.

⁴⁹⁹ See Kitcher 1990, pp. 11-14; Hoppe 1983, pp. 19-20; and Brook 1994.

and demonstrated science, because, like applied logic, it depends on empirical and psychological principles. (CPR, A 54-5/B 78-9)

Here Kant claims: (1) Pure ethics gives 'necessary laws' to the will, and the doctrine of virtues (what Kant calls 'anthropology' in the Groundwork) examines the conditions that 'hinder or help' compliance with these laws. (2) Applied logic, like the doctrine of virtues, investigates the conditions that may hinder or help compliance with the 'rules of the necessary employment' of the understanding. (3) Pure ethics stands to the doctrine of virtues *in the same relation as* pure general logic stands to applied logic. This line of thought very strongly suggests that the 'necessary rules' supplied by pure ethics have the same status as the 'necessary rules' supplied by pure logic. Since the former rules are certainly normative rather than empirical laws, it seems that the latter must be normative laws as well. I would challenge the proponent of a psychological interpretation of Kant's pure logic to show how this argument can be avoided.⁵⁰⁰

Does this textual argument also tell against Longuenesse's approach? Longuenesse, after all, does *not* agree with Kitcher that Kant's rules of synthesis are meant to be understood as empirical laws whose cognition gives us insight into how the mind works. Remember, however, that Longuenesse *does* agree with Kitcher that these rules are not to be understood as prescriptive normative laws. So *if* one thinks that pure ethics aims at establishing prescriptive norms that are meant to stake their claim from within the conscious perspective of practical deliberation, one should agree that Kant's suggestion that the laws established by pure ethics have the same status as those established by transcendental logic tells against Longuenesse's claim that transcendental logic is not aimed at establishing prescriptive norms that are meant to stake their claim from within the conscious perspective of the synthesizing subject. Now, Longuenesse might⁵⁰¹ respond by saying that pure ethics is *not* supposed to yield normative prescriptions either. She might say that pure ethics is only meant to give an analysis of a 'specific mode of volition', namely, of willing in accordance with a priori valid norms. This response raises many important issues, but it seems that, for instance, in the *Metaphysics of Morals* – which is part of pure ethics – Kant's ambition is to give an exhaustive catalogue of various types of obligations; and here it seems very compelling to understand him as intending to give explicit, systematic practical guidance.

Presently (in section 5) I will return to Longuenesse's important reasons for thinking that Kant's logic cannot aim at articulating prescriptive norms. For now, I want to suggest that the analogy between pure ethics and pure logic establishes a – admittedly inconclusive – presumption in favor of the idea that the rules of pure transcendental logic are intended as prescriptive norms; and I want to notice some implications of such an interpretation. First, I want to emphasize that I reject the idea that Kant's Analytic cannot be legitimately concerned with

⁵⁰⁰ Brook makes the remarkably implausible attempt to meet this challenge by arguing that even the laws of ethics are empirical: "these laws are not nonempirical, in the sense that we could not observe them in particular virtues. On the contrary, they are what is common to all virtues" (1994, pp. 6-7). This remark is confused in several ways. First of all, Kant holds that we *cannot observe* the influence of the moral law in actions that occur in the empirical world; if we could, then we would be able to know, by inference from observable actions, the moral character of agents, a possibility that Kant strictly denies. Second, virtue, for Kant, is the capacity to comply with the moral law as a result of struggle against the influence of sensible inclination; so it is obscure what it means to say that moral laws are 'common to all virtues'. But most importantly, it is abundantly clear (cf. CPR, A 548-9/B 576-7) that moral laws are meant to be categorically different from empirical laws: they say that actions are necessary "even although they have never taken place, and perhaps never will take place".

⁵⁰¹ My attribution of this line of thought to Longuenesse is based on personal correspondence.

mental *activity*.⁵⁰² Here the analogy between 'pure logic' and 'pure ethics' is, once more, instructive. Moral laws, for Kant, do not give us insight into *how the will works*, but they are intended to have *actions* as their effects.⁵⁰³ Likewise, while the pure rules of the understanding are not aimed at describing the mind, they are intended to have mental effects: namely, a proper combination of intuitions that yields cognition of objects.⁵⁰⁴ Hence, the undeniable fact that in Kant's Analytic the notion of synthesis qua mental activity is absolutely central need not tell in favor of a psychological or naturalistic interpretation; and conversely, the rejection of the psychological reading need not force us to ignore or dismiss Kant's appeal to mental activity.⁵⁰⁵

Second, I want to note that if we interpret Kant's rules of synthesis as prescriptive rules, his commitment to the idea that his pure logic is concerned with the unknowable noumenal mind makes perfect sense. It implies that Kant is occupying the normative standpoint, a perspective from which one aims at establishing rules that express what ought to happen: how we ought to will (pure ethics), and how we ought to think (pure logic). From this standpoint, the object of reasoning – the addressee of norms – must be represented as spontaneous, and hence as standing outside the deterministic natural order: this is why an investigation of how we ought to think takes as its object the noumenal mind (see my chapter VIII, section 4). That we cannot understand how the noumenal mind works is no objection to what Kant is doing: a normative investigation of how the mind should think does not aim at knowledge of how the mind works. This interpretation, if defensible, would put Kant's investigation of the rules that govern the synthesis of intuitions on a par with his non-descriptive investigation *both* of practical laws for acts of volition *and* of epistemic laws for reflective acts of concept-formation.⁵⁰⁶

3: Blind and conscious syntheses

One important reason that Longuenesse and Kitcher adduce against the normative account of

⁵⁰² Strawson 1966 (pp. 15-6) argues that Kant's emphasis on mental activity is wholly confused, and that one should focus on the 'legitimate, analytic' parts of the Analytic (i.e., what Strawson reconstructs as transcendental arguments). The tendency to focus on the 'epistemological' aspects of Kant's Analytic and bracket his talk about mental activity is widespread among commentators with very diverse commitments; one can witness it in Allison 2004 as well as in Guyer 1987 (see pp. 303-5; p. 315). See Anderson 2001 for a detailed discussion of the contrast between, and the respective shortcomings of, psychological and 'merely logical' readings of the Analytic.

⁵⁰³ Practical reason 'expects empirical effects from its ideas' (CPR, A 548/B 576).

⁵⁰⁴ Kitcher often appeals to the fact that Kant repeatedly speaks of *actions* of the understanding in his Analytic; and she complains that anti-psychological interpretations try to make sense of this talk in terms of the rather obscure notion of a 'logical act' (for this tendency, see Allison 2004 and Paton 1970). Kitcher asks (1993, p. 5): "how can something be part of an actual process and yet be merely logical?" I agree that the language of 'logical acts' is not helpful here, but I reject the idea that Kant's use of the concept of action commits him to describing or predicting activity. Kant's *practical* philosophy is full of references to mental acts of practical legislation or determining one's will, and here Kant is clearly not engaged in describing or observing the workings of the will (consider, for instance, his famous idea that we must 'act under the idea of freedom').

⁵⁰⁵ I take this point to be congenial to the reading offered by Anderson 2001, who also seeks to strike a balance between psychological and 'merely (epistemo)logical' interpretations of the Analytic by understanding rules for synthesis as normative rules (see, for instance, p. 277). See, however, my footnote 493 above.

⁵⁰⁶ The laws for reflectively analyzing intuitions into empirical concepts include the laws that govern the reflective power of judgment. For a clear statement that Kant considers these laws to be normative rather than empirical, see KU, 5: 182 (compare EEUK, 20: 239; see my chapter VIII, section 3). In an earlier essay (1991), Kitcher acknowledges that at least Kant's regulative epistemic principles have the status of normative laws that do not specify how we in fact do think; but in a more recent paper (1999) she treats even these rules in functionalist terms as resting on natural, mental dispositions that determine how we in fact think (p. 429).

rules of synthesis that I am envisaging is their conviction that the combination of sensible intuitions that brings them to synthetic unity is a blind, unconscious process. Hence, it cannot be governed by prescriptive norms. I agree that a process that takes place at a 'sub-personal' level cannot be governed by norms. So the question is whether we should read Kant as holding that the objective synthesis of sensible intuitions is a blind, non-conscious mental activity.

Kant clearly thinks that synthesis, in *some*, even in a general sense, is a blind, unconscious process. In a famous passage (CPR, A 78/B 103) he says that "synthesis as such...is the mere result of the power of imagination, a blind but indispensable function of the soul, without which we should have no knowledge whatsoever, but of which we are scarcely ever conscious." Note, however, that he adds: "[Allein,] To bring this synthesis [the 'blind function of the soul'] to concepts is a function which belongs to the understanding, and it is through this function of the understanding that we first obtain cognition properly so called." The connective 'Allein' signals a contrast, and hence the rhetorical upshot of this passage seems to be that there are certain types of synthesis that are blind and unconscious whereas others – those that bring the (result of the) former synthesis to concepts – are not.⁵⁰⁷ Now, the notion of 'bringing representations to concepts' is ambiguous: it might refer to the formation of empirical concepts *or* to the act that brings given intuitions 'under' the categories and thereby to synthetic unity.

Before reviewing whether the synthesis of intuitions that brings them to objective unity (and under the categories) might qualify as belonging to the second, conscious type of synthesis that Kant refers to at A 78/B 103, I first want to point out that Kant does not think that *every* kind of synthesis is relevant to his transcendental investigation in the *Analytic*. In the most psychologically sounding passages of the *Analytic*, which can be found in the A-Deduction, Kant distinguishes, among other things, between a synthesis of reproduction that is performed *by the imagination* and a synthesis of recognition *in concepts* (a contrast remarkably similar to the one introduced at A 78/B 103). Now, about the former synthesis, Kant says:

It is a merely empirical law, that representations which have often followed or accompanied one another finally become associated... (CPR, A 100) **In so far as imagination is spontaneity**, I sometimes also entitle it the productive imagination, to distinguish it from the reproductive imagination, whose synthesis is entirely subject to empirical laws, the laws, namely, of association, and which therefore contributes nothing to the explanation of the possibility of *a priori* knowledge. The reproductive synthesis falls within the domain, not of transcendental philosophy, but of psychology (CPR, B 151-2; my emphasis)

These passages have a number of important implications. First, Kant thinks that a synthesis that is 'entirely subject to empirical laws' is *not* to be regarded as spontaneous (of course, this is not at all surprising if one keeps in mind Kant's general definition of spontaneous action as causally undetermined).⁵⁰⁸ Second, it is precisely the kind of non-spontaneous synthesis that is governed

⁵⁰⁷ For a similar reading of this passage, see Anderson 2001, pp. 80-1. See also Allison 2004, pp. 170-1.

⁵⁰⁸ This is worth emphasizing because there is a tendency among commentators to suppose that it is a sufficient condition on the 'spontaneity of the mind' that the products of this activity are not determined by the objects that affect our senses (or by the perceptual data they affect us with); the mind's spontaneity is said to be compatible with empirical determination by 'inner', psychological laws and causes. This deflationary conception of epistemic spontaneity is at work in Kitcher (I find it quite ironic that she complains about McDowell that he offers "an anodyne... reading of 'spontaneous' as 'conceptual'"; 1999, p. 424). See also Dickerson 2007, who attributes to Kant the worry (p. 44, pp. 53-55, p. 60) that the categories might have their origin *in spontaneity*, namely the Humean laws of reproductive imagination, and yet lack objective validity. As we saw in chapter IX, Sellars 1970 also endorses a deflationary conception of the ('merely relative') spontaneity of mind. Allison 1996, p. 57

by empirical laws that Kant deems irrelevant to transcendental logic. Finally, we can here see that my claim that Kant is not doing psychology in the *Analytic* is perfectly compatible with the concession that he draws on (what he thinks are) established, general psychological facts about how the empirical mind operates in formulating his rules for thought. Here the method of pure logic is, again, not in principle distinct from that of pure ethics: the former articulates its rules for thought against the background of the general psychological idea that we are passively affected with sensible intuitions that are combined according to associative psychological laws; and the latter articulates its rules for volition against the background of the general psychological idea that we are passively affected with empirical desires that represent their object as pleasant.

Now, the most important question for our purposes is whether the synthesis of given intuitions that relates them to objects, that is spontaneous, and that is directed by the categories, is a blind or rather a conscious process. And here I want to suggest that there is significant textual evidence for the assumption that Kant thinks that this kind of synthesis, *in contrast to reproductive empirical synthesis*, does occur at the conscious level:

This thoroughgoing identity of the apperception of a manifold which is given in intuition contains a synthesis of representations, **and is possible only through the consciousness of this synthesis**. For the empirical consciousness, which accompanies different representations, is in itself diverse and without relation to the identity of the subject. That relation comes about, not simply through my accompanying each representation with consciousness, **but only** in so far as I conjoin one representation with another, **and am conscious of the synthesis of them**. (CPR, B 133; my emphasis)

The transcendental unity of apperception is that unity through which all the manifold given in an intuition is united in a concept of the object. It is therefore entitled objective, and must be distinguished from the subjective unity of consciousness, which is a determination of inner sense – through which the manifold of intuition for such [objective] combination is empirically given. Whether I can become empirically conscious of the manifold as simultaneous or as successive depends on circumstances or empirical conditions. Therefore the empirical unity of consciousness, through association of representations, itself concerns an appearance, and is wholly contingent. (...) Only the original unity is objectively valid; the empirical unity of apperception, upon which we are not here dwelling...has only subjective validity. To one man, for instance, a certain word suggests one thing, to another some other thing; the unity of consciousness in that which is empirical is not, as regards what is given, necessarily and universally valid. (CPR, B 139-40)

Here Kant draws a contrast between two kinds of synthesis which maps onto a contrast between two kinds of consciousness, i.e., two forms of being aware of one's representations. The contrast is between an empirical, 'subjectively valid' and a transcendental, 'objectively valid' unity of intuitions in one consciousness. The latter unity is the synthetic unity of intuitions in which sensible representations are related to objects (i.e., subsumed under the categories). Now the crucial point is this: as the first passage intimates, Kant thinks that it is a *sufficient condition* for the empirical unity of consciousness that I am conscious of each of my various sensible representations; whereas it is a *necessary condition* for the transcendental unity of consciousness that I am *conscious of the synthesis of* the representations that brings them under the categories and hence relates them to objects. As the second passage shows, the empirical unity of consciousness is itself the result of synthesis, namely, of the reproductive synthesis that stands under associative, empirical laws and hence depends on empirical conditions that may vary from subject to subject. With regards to this synthesis, we are aware only of its *result* (namely, of the

affirms that the understanding is absolutely spontaneous – however, he thinks that this affirmation is devoid of any ontological implications. I have discussed this rather puzzling interpretation at length in chapter II, section 3.

fact that certain types of representations are, in our mind, associated with others), *but not of the synthesis itself*: the combinatory processes that glue our representations together so that they eventually prompt each other in our conscious mind are indeed blind and 'governed' by rules only as the movement of the planets is governed by the law of gravity.⁵⁰⁹

Kant's idea that the spontaneous synthesis of given intuitions is a *conscious* mental act can also be found in the A-Deduction.⁵¹⁰ So there is what I take to be undeniable textual evidence for interpreting Kant as holding that the significance of being conscious of one's mental activity is not limited to the formation of conceptually determinate judgments but also extends to the original synthesis of given intuitions that relates these intuitions to public empirical objects.⁵¹¹

Now, the above passages also exemplify a very important *stylistic* point: namely, Kant's use of the *first-person present tense*. He says, for instance, that the objective unity of given intuitions comes about "only in so far as I conjoin one representation with another, and am conscious of the synthesis of them". This is an obvious problem for Kitcher's idea that Kant "doubts that we [as subjects of synthesis] are aware" of the synthesis of intuitions. Her 'solution' consists in accusing Kant of (yet another) incoherence, namely, of making "unwitting shifts

⁵⁰⁹ Longuenesse seems to think (2001, pp. 42-3) that even empirical, associative combination of representations is guided by the goal to represent a whole of experience. I do not see how a blind process that is completely removed from the subject's reflective awareness can be considered goal-directed. However, I think that her view that synthesis is directed at the goal of representing a unified course of experience is a very illuminating formulation of the point and purpose of conscious, spontaneous synthesis.

⁵¹⁰ See CPR, A 103-4 and especially A 108. In the latter passage, Kant is clearly concerned with the original synthesis that binds given intuitions together in *one* consciousness and that relates them to objects; and he thinks that this requires consciousness of the identity of the *act* of synthesis: "This transcendental unity of apperception forms out of all possible appearances, which can stand alongside one another in one experience, a connection of all these representations according to laws. For this unity of consciousness would be impossible if the mind in knowledge of the manifold could not become conscious of the identity of function whereby it synthetically combines it in one knowledge. The original and necessary consciousness of the identity of the self is thus at the same time **a consciousness of an equally necessary unity of the synthesis of all appearances according to concepts**, that is, according to rules.... For the mind could never think its identity in the manifoldness of its representations, and indeed think this identity a priori, **if it did not have before its eyes the identity of its act, whereby it subordinates all synthesis of apprehension (which is empirical) to a transcendental unity**, thereby rendering possible their interconnection according to a priori rules." (My emphasis.)

One may object to my reading of these passages (as well as those quoted from the B-edition) that they show only that Kant is committed to a 'conditional' awareness of synthesis: insofar as I bring my representations to the transcendental unity of consciousness, I must be conscious of the act of synthesis. This does not establish the unconditional awareness of synthesis of all my representations. I accept this point, but I think that this is harmless for my interpretation. For I am not claiming that I am aware of a synthesis with regard to all my representations. Rather, and in accordance with what Kant says early on in the B-Deduction, I am only claiming that every representation I can call mine must be a possible component of a conscious synthesis.

⁵¹¹ Wolff 1963 (see especially pp. 130-1) claims that if Kant were to allow that the act of combination that brings given representations to synthetic unity is a conscious process, he would be caught in a vicious circle because synthesizing activity is "what...makes [the unity of] consciousness possible" in the first place. This objection presupposes that Kant's account of synthesis is meant to explain how a unified consciousness arises *absolutely*. But Kant's ambition is not to explain how a unified consciousness first 'pops up' from nowhere (from a "transcendental limbo", as Wolff puts it). Rather, he articulates normative conditions for achieving a unified state of consciousness *with respect to some given representation(s)*: his conception of (transcendental) synthesis concerns the act of bringing a given representation, or set of representations, to the unity of consciousness, i.e., of integrating representations into a unified whole of experiences that I consider 'mine'. Engaging in this act presupposes that one already has a unified whole of conscious states at one's disposal, to which a given experience can be considered as 'belonging'.

between the perspective of the individual who is engaging in various mental activities and that of the theorist who is describing those activities".⁵¹² But if we give up the idea that the spontaneous synthesis of intuitions is a sub-personal, pre-conscious process, there is nothing 'unwitting' in Kant's methodology; rather, we can conjecture that his use of the first-person present tense is to be explained by the fact that Kant qua theorist finds it indispensable, for the purposes of expounding the point and the character of rule-governed synthesis, to appeal to the perspective of the synthesizing subject and to capture the 'phenomenology' of bringing objective order to the given, arbitrarily combined manifold of intuitions. I will return to this issue in section 5. But I want to note that there is an important reason why Kant should deem the first-person awareness of rule-governed synthesis indispensable. The point is that for Kant, the very notion of the unity of the self, or self-identity with respect to various different representations, collapses into the notion of the unity or self-identity of the *act* of combining various representations. This is his attempt to save the legitimacy of the concept of personal identity against Hume's attack without giving in to the Cartesian idea that personal identity involves awareness of a thinking substance. But this means that the notion of personal identity, for Kant, has significance only and exclusively from the first-person perspective of the agent of synthesis.

4: The rules of synthesis

I argued that there is a strong textual basis for attributing to Kant the view that the objective synthesis of intuitions is a conscious mental activity. Conscious synthesis can be governed by reflective awareness of normative rules. I now want to ask what these rules are.

We saw that the objective unity that results from relating intuitions to objects is different from the subjective unity that results from our sensible representations becoming associated with one another. Kant thinks that the kind of synthesis that achieves objective unity *operates on* the results established by the empirical synthesis that gives associative unity to representations: "the subjective unity of consciousness... is a determination of inner sense – through which the manifold of intuition for such [objective] combination is empirically given" (CPR, B 139-40). Kant specifies further that

...whether I can become empirically conscious of the manifold as simultaneous or as successive depends on circumstances or empirical conditions. Therefore the empirical unity of consciousness, through association of representations...is wholly contingent.

So in the Deduction Kant suggests that the original synthesis of intuitions transforms a merely empirical consciousness of representations into an objective unity of consciousness; and he uses *temporal* predicates to characterize these two kinds of unities. Now, the representation of an objective temporal order is enabled by the relational categories: substance, causality, and community. In what follows I shall focus on synthesis under these concepts. This synthesis requires, as we saw (CPR, B 133), that the subject be conscious of the *act* of synthesis.⁵¹³

⁵¹² Kitcher 1993, p. 83.

⁵¹³ The 'blind' account of synthesis is more plausible with respect to the synthesis that stands under the *mathematical* categories. This synthesis brings intuitions only to the *subjective, empirical* unity of consciousness (CPR, A 162/B 202; B 207), and it is directed at perception rather than experience (CPR, A180/B 223). Kant frequently appeal to a conscious awareness of the mathematical synthesis of the manifold (CPR, A 103-4; A 162/B 203), but the proponent of the sub-personal account might be able to deflate this by holding that the consciousness must pertain only to the *product* rather than to the *act* of synthesis. I do not want to take a definite

I shall argue that the rules that govern the synthesis that brings intuitions to objective temporal unity are the Analogies of Experience. I will begin here with the observation that the picture sketched by Kant in the Deduction, according to which subjectively or arbitrarily reproduced representations provide the material for the synthesis that brings given intuitions to objective temporal unity, is very much in play in the passages where Kant expounds the rules he calls Analogies of Experience. The 'general principle' of the Analogies is that the cognitively significant state Kant calls 'experience' is possible only through a necessary combination of what Kant calls perceptions (CPR, B 218). This necessary combination yields precisely the kind of objective *temporal* unity of given representations that Kant is concerned with in the Deduction (CPR, B 139-40), namely (CPR, A 177/B 220) the "synthetic unity in the time-relations of all perceptions" "to which everything that is to belong to my cognition, and so can be an object for me, has to conform". Having perceptions is by itself not sufficient for experience: experience "is a synthesis of perceptions, not contained in perception but itself containing in one consciousness the synthetic unity of the manifold of perceptions". Hence, the synthesis that yields a necessary, non-arbitrary connection of given representations *operates on* perceptual states that are already temporally connected, albeit only in an arbitrary and subjective fashion.⁵¹⁴ Now, since the 'general principle' of the Analogies is that experience requires a necessary connection of perceptions, we can conjecture that the Analogies must be the rules that guide the transition from an empirical unity of perceptions to experience, i.e., from a subjective order of representations in an empirical consciousness to the necessary connection of perceptions in an objective (transcendental) unity of consciousness.⁵¹⁵

stand on this issue here. One might object to my argument as a whole that it drives a wedge between synthesis according to the mathematical and synthesis according to the dynamical (especially relational) categories. In response, I want to say that (i) 'mathematical' and (ii) 'dynamical' synthesis *must* be understood in very different ways. For once, as I already mentioned, it is *only* (ii) that brings intuitions to the objective, transcendental unity of consciousness. But it is also very hard to even state clearly (a) what the rules for synthesis corresponding to the mathematical categories are supposed to be and (b) what it would take to violate these rules. I find it very telling that Anderson (2001, p. 292) gives the following example: "we would dismiss anyone who calculated the area of our elliptical table top by some other rule than p times the product of the radii"; likewise, Guyer 1987, pp. 195-6 argues that the principles associated with the mathematical categories are not concerned with the phenomenology of perception but with the theory of measurement. But the appeal to mathematical (mis)calculation of a figure in space can hardly illuminate how we should understand the rule-governed synthesis of intuitions that allows for the representation of objects in space in the first place. In fairness to Anderson and Guyer, Kant himself very much suggests that rules for mathematical synthesis are concerned with calculation or measurement (CPR, A 178-9/B 221); but that makes it all the more obscure how he conceived of the kind of mathematical synthesis that must be epistemically prior to any attempt to carry out mathematical calculations.

⁵¹⁴ CPR, B 219: "...perceptions come together only in accidental order, so that no necessity determining their connection is or can be revealed in the perceptions themselves". Kant considers it a psychological fact that our given representations are always successive; but this does not determine any *objective* temporal relations among the objects of perception. The successive character of our perceptual course is insensitive to whether what is being perceived is simultaneous or successive. The order of time in which the apprehension of the parts of a house occurs is successive, but these parts exist simultaneously; I could have perceived them in any order. On the other hand, the order of time in which I perceive the course of a ship driving down a stream is also successive, but here the temporal order of perceptions corresponds to the course of the ship (CPR, A 191-3/B 236-8).

⁵¹⁵ The appeal to such a transition raises very difficult interpretive and philosophical questions. Does Kant think that in order to qualify as a thinker who combines given perceptions into an objective unity of consciousness, one must be conscious of the material to-be-combined as merely subjective? Wouldn't that commit him to the idea, deemed unattractive by many post-Sellarsian philosophers, that cognizers must work their way up from a pre-

I want to confirm this assumption by looking at two passages in which Kant designates the overall point of the rules he calls Analogies of experience. Here is the first:

This synthetic unity in the time-relations of all perceptions, as thus determined a priori, is the law, that all empirical time-determinations must stand under rules of universal time-determination. The analogies of experience, with which we are now to deal, must be rules of this kind. These principles have this peculiarity, that they are not concerned with appearances and the synthesis of their empirical intuition, but only with the existence of such appearances and their relation to one another in respect of their existence. (CPR, A 177-8/B 320)

Here Kant explicitly refers to the Analogies as rules of time determination *that are concerned with* the synthesis that brings an objective order into our representation of the temporal relations among appearances. The last sentence of the above passage means that the Analogies qua rules for synthesis concern not the synthesis of *one* empirical intuition (mathematical synthesis), but the synthesis that relates various separate intuitions to one another in a way that implies existential dependence relations among appearances (cf. CPR, A 179/B 222).

The second passage (CPR, A 181/B 223-4) is even more explicit:

...the principles [of the understanding, including the Analogies] can therefore have no other goal [Ziel] save that of being the conditions of the unity of empirical knowledge in the synthesis of appearances.

Here Kant refers to the *goal* of the Analogies: this goal is the achievement of a unity through (in) synthesis. If it is the sole purpose of the Analogies to function as conditions of unity in synthesis, it seems to follow that the Analogies are rules that govern synthesis.

Now, one might object that Kant cannot have in mind here the kind of synthesis that brings given intuitions to the objective unity of apperception, because (as we saw) this synthesis is prior to determinate empirical thought, whereas the synthesis in the above passage is said to be a condition of the unity of *empirical knowledge*; such knowledge, one might say, requires the synthesis of empirical concepts in fully determinate judgments of experience. Here we need to get clear about what Kant means by 'empirische Erkenntnis'. This term takes on a number of different connotations in Kant, but one sense that is operative in the Principles chapter is explicitly defined in the very first sentence of the 'proof' of the Analogies (CPR, B 218): empirical knowledge is said to be "a knowledge which determines an object through perceptions". The 'determination of an object' for given representations is repeatedly specified by Kant as *the* point of the synthesis that precedes the 'analysis' of intuitions into empirical

conceptual 'given' – an idea that, as some suggest (Abela 2002; see footnote 443 above) is wholly alien to Kant? Various textual data suggest that Kant indeed thinks that we have such pre-objective consciousness of our perceptions (for a systematic defense of this idea, see Prauss 1971). First, this is suggested by the very notion of a merely subjective, empirical unity of consciousness. Second, of course, there is the appeal to subjective judgments of perception in the Prolegomena. Third, Kant defines perceptions, which in themselves are only subjectively combined, as conscious states (CPR, A 163/B 203; B 208; A 320/B 377); and he states in very unambiguous terms: "I am conscious only that my imagination sets the one state before and the other after, not that the one state precedes the other in the object. In other words, the objective relation of appearances that follow upon one another is not to be determined through mere perception" (CPR, B 233-4). Two points can be made to mitigate these remarks. First, perceptions as such are not entirely pre-conceptual but already structured by the mathematical categories (cf. footnote 513 above). Second, Kant might mean only that we may *potentially* become conscious of merely subjective perceptual states as the material for objectifying synthesis, and that what figures in our actual conscious awareness is, typically, our awareness of the determination of our thoughts by rules of synthesis (but see Prauss 1971 for systematic reasons that tell against this conjecture).

concepts; and this synthesis is the focus of the enquiry in the Deduction and the Principles chapters.⁵¹⁶ Hence, there is every reason to think that Kant uses the term 'empirical cognition' here *not* to designate the kind of cognition that we get when we form and apply (valid) empirical concepts, but to pick out a more fundamental kind of thought: namely, a representation that qualifies as empirical *Erkenntnis* in so far as it brings given intuitions under concepts of *empirical objects in the most general sense*, namely, under the (schematized) categories, which are "fundamental concepts by which we think objects in general for appearances" (CPR, A 111). That is, by applying these concepts to appearances we succeed in representing the empirical character that objects of experience must have merely *as such*, regardless of how this empirical character is *specifically* to be understood according to further determination of objects through empirical concepts. For instance, we succeed in representing something as a substance that persists through change, where this leaves indeterminate *what kind of* substance we are dealing with. Determining species of empirical substances requires empirical concepts, and it is by forming such concepts that we progress from the representation of empirical objects *in general* to systematic empirical cognition of specific substances, powers, and causal laws.⁵¹⁷

A second worry about my claim that the Analogies are subject-directed rules for synthesis is that the Analogies seem to be *descriptive claims about objects*. Take the Second Analogy in the A-edition: "Everything that happens...presupposes something upon which it follows according to a rule" (CPR, A 189). How can this statement possibly be conceived as a *rule that governs our mental activity*? My response is that the Analogies owe their status as valid principles about the way objects must be to their status as rules that specify necessary conditions for experience of objects. To explain this idea I shall focus on the Second Analogy (the so-called causal principle).

The basic idea that informs the Second Analogy is that an alteration, such as the change of a substance from state *s1* to *s2*, can be represented only if an objective temporal course can be represented in which the original state (*s1*) is temporally prior to the resultant state (*s2*). Time cannot be directly perceived, and passively received perceptions always occur successively,

⁵¹⁶ In the Deduction, see CPR, B 138: "The synthetic unity of consciousness is, therefore, an objective condition of all knowledge. It is not merely a condition that I myself require in knowing an object, but is a condition under which every intuition must stand **in order to become an object for me**. For otherwise, in the absence of this synthesis, the manifold would not be united in one consciousness" (my emphasis; see also CPR, A 108-10). The same kind of unity is at issue in the Principles chapter (CPR, B 220): "The general principle of the three analogies rests on the necessary unity of apperception, in respect of all possible empirical consciousness...and since this unity lies a priori at the foundation of empirical consciousness, it follows that the above principle rests on the synthetic unity of all appearances as regards their relation in time. For the original apperception stands in relation to inner sense...and indeed a priori to its form, that is, to the time-order of the manifold empirical consciousness. All this manifold must, as regards its time-relations, be united in the original apperception. This is demanded by the a priori transcendental unity of apperception, to which everything that is to belong to my knowledge, **and so can be an object for me**, has to conform" (my emphasis).

⁵¹⁷ Compare the following passage at CPR, A 201-2/B 246-7: if there is "an order of successive synthesis that determines an object", i.e., if "my perception is to contain knowledge of an event, of something as actually happening", then "it [the perception] must be an empirical judgment in which we think the sequence as determined; that is, it presupposes another appearance in time, upon which it follows necessarily, according to a rule." The 'empirical judgment' here consists in the thought that the perceived succession must be due to another event from which it follows according to a necessitating empirical law. It is through forming this thought that the succession becomes an empirical object *in general* for me. My representation of the succession remains at this general level until I can cognize *which* event causes it and *what* laws govern it. For the idea that the subsumption of intuitions under the categories, which is epistemically prior to the application of empirical concepts, qualifies as a judgment 'in general', see Prol, 4: 300 and Man, 4: 475; I discuss this point in chapter XI, section 2.

regardless of whether the states they represent are successive or simultaneous. Hence the subject has to combine its perceptions in such a way that it can regard s_1 as objectively prior to s_2 . This, Kant thinks, is possible only if one regards the change from s_1 to s_2 as following necessarily from a preceding event according to a rule (CPR, A 200/B 245). The concept of causality (uniquely) allows for the representation of a change that necessarily occurs according to a rule. The use of this concept is thus a necessary condition for representing an alteration. Now, Kant *identifies* a condition on the possibility that our thought may give us experience of objects with a condition on the character of the objects thus experienced: "the conditions of the possibility of experience...are at the same time conditions of the possibility of the objects of experience" (CPR, A 158/B 197). This identification provides us with an answer to the question of how the object-directed principle (1), 'Everything that happens...presupposes something upon which it follows according to a rule' could be understood as a subject-directed rule for the combination of our sensory representations. The *very same principle* could also be phrased as follows (2): 'The representation of an alteration requires that one combines one's discrete perceptions of two states s_1 and s_2 of a substance by positing ("setzen" is Kant's term) some causal ground that necessarily and invariably brings about alterations of the states of substances from s_1 to s_2 .' For Kant, (1) and (2) are, in a way, equivalent, although the sense in which they are cannot be captured in terms of extensional or intensional equivalence. The equivalence derives from the notion that objects of knowledge depend for their character on our cognitive faculties. We might put this point by saying that the principles that elucidate necessary conditions for the experience of objects are *ontologically transparent* to the way objects must be. Kant thinks that if this idea of ontological transparency is given up, and if it is assumed that the properties of objects of knowledge are independent of our cognitive faculties, there is absolutely no way of vindicating any principles that purport to give us necessary, non-trivial truths about objects of knowledge.⁵¹⁸

⁵¹⁸ See CPR, BXVI-BXVIII. Kant's worry about the compatibility of transcendental realism and synthetic *a priori* knowledge, as I understand it, is this: we are convinced that *every* alteration has a cause, or that there *must* be a reason for why events happen as they do. These claims express our commitment to the idea that nature as such is necessarily structured in an intelligible manner. Hence, the causal principle is intended to apply to natural events with strict necessity and universal validity. But how can I know that a non-trivial principle like this really does apply to all natural events? If objects of knowledge are supposed to be entirely independent of our cognitive faculties, then, Kant thinks, there is no way of ensuring that objects *must* be a certain way (apart from uninformative points such as, 'an alteration presupposes something that precedes in time'). Now, Kant thinks that we *can* demonstrate in what way we must exercise our faculties in order to cognize objects – hence we can also demonstrate what objects of knowledge must be like *if* we give up the realist assumption that the ontological constitution of objects of knowledge is independent of our cognitive faculties. The *modal* aspect of the knowledge Kant is concerned with is suppressed in the criticisms of Strawson 1966, p. 244 and Walker 1978, pp. 11-13. By contrast, Guyer 1987, pp. 349-69 is very sensitive to the idea that Kant's idealism is motivated, to a considerable extent, by the aim to make room for knowledge of *de re* necessities. But the status of his criticisms of Kant's enterprise is a bit obscure. Guyer holds that Kant can establish *only* principles of the form (P), 'If we are to perceive objects, we must represent them in certain ways (e.g., as satisfying Euclidean principles)'. He argues that Kant has no argument for the idealist conclusion that the objects of our knowledge must answer to these subjective conditions on representation. Now, at one point (p. 349), Guyer seems to suggest that principles like (P) do allow for a priori knowledge of objects because the necessity that our experience must be a certain way is "evidence" for the way objects must be. This is fallacious: that we must think about objects in a certain way provides absolutely no evidence for the way mind-independent objects must be. (I think that it doesn't even provide evidence for the way objects actually are; see footnote 539 below.) Later in his discussion (see pp. 366-9), Guyer seems to concede that (P) does not imply any a priori knowledge about objects, but then I don't understand what his criticism is supposed to amount to. It cannot be, as Guyer seems to suggest on p. 368, that if

To further support this interpretation I now want to present the full context of the above passage where Kant specifies the 'goal' of the Analogies (CPR, A 181/B 223-4):

For if the objects to which these principles [the Analogies] are to be related were things in themselves, it would be altogether impossible to know anything of them synthetically a priori. They are, however, nothing but appearances; and complete knowledge of them, in the furtherance of which the sole function of a priori principles must ultimately consist, is simply our possible experience of them. The principles can therefore have no other goal [Ziel] save that of being the conditions of the unity of empirical knowledge in the synthesis of appearances.

Here Kant first expresses his general point that we can only establish non-trivial, necessary claims about how objects must be if we give up the idea that these objects are independent of our cognitive faculties. *On this basis*, Kant goes on to say that the (sole) goal of the Analogies is that of being conditions *of a synthesis* that provides us with experience of objects. This, I think, illustrates my claim that Kant identifies rules of synthesis with principles about the way objects must be. For Kant, the conditions or rules that govern the synthesis that relate intuitions to objects, and that make objective experience possible, are "at the same time" conditions for the possibility of objects of experience, and because of this congruence there are objectively valid synthetic a priori principles that articulate conditions on both experience and on objects.

I now want to address a competing account of the rules that govern the synthesis that aims at bringing our representations to objective temporal unity. Some commentators argue that the synthesis that leads to temporal unity of representations is guided not by the Analogies but by the *schemata* of the relational categories.⁵¹⁹ Here I must begin by confessing that it is very much unclear to me how a predicate such as 'succession' (the schema of the category of causality) could *all by itself* govern, as a rule, the synthesis of perceptions that relates them to objects.⁵²⁰ I want to argue that according to Kant, the triad of category, schema and Principle is separable only in theory, but not in application (to appearances). Consider first what Kant says immediately prior to his derivation of the system of Principles at CPR. A 160-2/B 199-201:

...in the application of pure concepts of understanding to possible experience, the employment of their synthesis is either mathematical or dynamical. (...) The table of categories is quite naturally our guide in the construction of the table of principles. For the latter are simply rules for the objective employment of the former.

We apply the categories to possible experience by employing them in synthesis; and we can derive the table of the principles from the table of the categories because the principles are "rules for the employment of" the categories. Since Kant believes that the employment of the categories lies in synthesis, this line of thought would make little sense *unless* he also believed that the principles are "rules for the employment" of the categories *in synthesis* (i.e., in the necessary

Kant assumes knowledge of de re necessities from the start, he circularly assumes transcendental idealism from the start: clearly, Kant's argument is that we can argue for idealism by showing that we can lay claim to knowledge of de re necessities *only* by assuming idealism – a condition that Guyer himself seems to accept.

⁵¹⁹ See Longuenesse 2001, p. 253, pp. 368-71.

⁵²⁰ My failure to understand this idea might be related to my difficulties to get a grip on Longuenesse's idea that the categorial schemata *guide the very acts of synthesis that produce them*: "they would not come forth as rules of synthesis unless synthesis according to these rules were occasioned by sensible impressions" (2001, p. 253). I see how a rule that we have already formed can govern an act of thought implicitly, without being in the foreground of thought. But I fail to see how a rule that has not at all 'come forth' can govern the very act of (blind) thought that makes it come forth. What would such 'governance' amount to?

combination of given intuitions).⁵²¹ Unsurprisingly, it turns out that the principles divide into two kinds (mathematical and dynamical) which correspond exactly to the two kinds of synthesis in which the categories are being employed. Now, Kant stresses that the principles are *objective, valid* rules for employing the categories, and it is this emphasis that requires us to take into account the schemata. Consider again the crucial passage at CPR, A 180-1/B 223-4:

...what has been said of all principles that are synthetic must be specially emphasized, namely, that these Analogies have significance and validity only as principles of the empirical...employment of understanding; that only as such can they be established; and that appearances have therefore to be subsumed, not simply under the categories, but under their schemata. (...) The principles can therefore have no other purpose save that of being the conditions of the unity of empirical knowledge in the synthesis of *appearances*. But such unity can be thought only in the schema of the pure concept of understanding. The category expresses a function which is restricted by no sensible condition and contains the unity of this schema, as of a synthesis as such. (...) In the principle itself we do indeed make use of the category, but in applying it to appearances we substitute for it its schema as the key to its employment, or rather set it alongside the category, as its restricting condition, and as being what may be called its formula.

Again, the picture is that we apply the categories to objects when we employ them in synthesis, and the principles are rules or conditions for such employment. But Kant adds that in applying the category to appearances, the schema imposes a constraint on the category to guarantee that both the category and the principle relate to objects only *qua* appearances. The general principle, 'Everything that happens...presupposes something upon which it follows according to a rule' yields a valid rule for synthesis that makes experience of objects possible, and hence (given the Copernican Revolution) a valid claim about objects, *only if* the concept of cause that is operative in this principle is restricted to appearances by the temporal predicate that yields the schema of causality. The schema of causality is *succession* of a real state in time to another state (CPR, A 145/B 184). Hence, the above principle yields a valid rule for synthesis and a valid corresponding claim about objects *only if* it relates to temporally situated objects, namely, to successions from one temporally determinate state of an object to a contrary and posterior state:

The principle of the causal relation in the sequence of appearances is therefore also valid of all objects of experience (under the conditions of succession), as being itself the ground of the possibility of such experience. (CPR, A 202/B 247) [Notice how this illustrates, once more, my idea that principles of the possibility of experience are 'ontologically transparent', i.e., are "at the same time" principles of the possibility of objects of experience]

So the relation between category, principle, and schema seems to be the following: the principle is the rule for the employment of the category in the synthesis of given intuitions that brings them to objective unity, and the schema both 'realizes' and 'restricts' (CPR, A 147/B 186) this employment by limiting it to temporal appearances. This restriction serves to guarantee the validity of the application of the rule that employs the category. Hence, while the categorial schemata undeniably *are* very important for the synthesis of sensible intuitions that brings them to objective unity, I do not believe that Kant thinks that these schemata are themselves the rules that govern this synthesis. Rather, the schemata add cognitive determinacy to and thereby restrict

⁵²¹ This synthesis must be the kind of synthesis that brings intuitions to synthetic unity: Kant is concerned here with the systematic derivation of all synthetic a priori laws of the understanding, and the highest common principles of these laws is (CPR, A 158/B 197) that "every object stands under the necessary conditions of synthetic unity of the manifold of intuition in a possible experience".

the scope of the principles qua rules of synthesis.⁵²²

Against this interpretation, Longuenesse might argue that Kant holds that the transcendental synthesis of the imagination produces the schemata of the categories *prior to* the emergence of the principles (and of the categories as – in her terms – 'clearly reflected' concepts). Kant does hold that the categorical schemata are produced by productive imagination, but I do not see clear textual evidence for the idea that according to Kant the categorical schemata are produced by a pre-conscious act of the mind that is prior to the recognition of the categories qua rules of synthesis. Consider the following, extremely suggestive but also elusive 'explanation':

...the schema of a pure concept of understanding...is simply the pure synthesis, **determined by a rule of that unity...to which the category gives expression**. It is a transcendental product of imagination, a product which concerns the determination of inner sense in general...in respect of all representations, so far as **these representations are to be connected** a priori in one concept in conformity with the unity of apperception. (CPR, A 142/B 181; my emphasis).

Here, the rule of pure (a priori, objective) synthesis is associated not with the schemata, but with the categories ("to which the category gives expression"). While it is impossible to draw anything definite from this passage, I want to suggest that the passage contains one very important clue for understanding Kant's conception of what the 'production' of the categorical schemata amounts to. Rather than reading Kant as holding that our awareness of the schemata is posterior to transcendental synthesis (i.e., emerges as the result of 'blind' synthesis), we could understand him as saying that our awareness of the schemata is *prior to* transcendental synthesis, in the sense that the production of the schemata is occasioned by an awareness of a manifold of representations as something *to be connected*. That is, Kant's point might be that our awareness of a manifold of given intuitions that is *to be brought* to synthetic unity (and hence *to be brought* under categories) affords us with a representation of the restricting sensible conditions (schemata) under which the rules for combining that manifold can be employed objectively.

In support of this reading, I want to notice that Kant explicitly holds that our awareness of transcendental (or 'figurative'; cf. CPR, B 151) synthesis and of its product (objective unity), *as something (yet) to be achieved*, precedes the actual performance of that synthesis and thereby provides the act of synthesis with a goal-directed character:

I am conscious to myself a priori of a necessary synthesis of representations -- to be entitled the original synthetic unity of apperception -- under which all representations that are given to me must stand, **but under which they have also first to be brought by means of a synthesis**. (CPR, B 135; my emphasis)

Moreover, there is explicit textual support for the idea that the *conscious awareness* of

⁵²² In this context it is significant to notice that Kant often simply identifies the categories with the principles. At CPR, B 167 he says that the categories must be considered "self-thought first principles a priori of our knowledge"; at CPR, B 168 he explains that "the Deduction is the exposition of the pure concepts of the understanding...as principles of the possibility of experience". At B 167, he adds that "the categories contain, on the side of the understanding, the grounds of the possibility of all experience as such. How they make experience possible, and what are the principles of the possibility of experience that they supply in their application to appearances, will be shown more fully in the following chapter" The 'following chapter' is, of course, the 'Analytic of Principles'. No doubt his identification of categories and principles rests on his thought that the latter are rules for making concrete use of the former in synthesis. Since the *objective* employment of the categories, in their application to *appearances*, requires the schemata as restricting conditions, the objective synthesis of intuitions that yields experience of objects requires the triad of categories, principles, and schemata.

transcendental synthesis and of the unity it aims at, as something *yet to be achieved*, informs us of the limiting conditions (schemata) under which this synthesis takes place. Kant considers transcendental or figurative synthesis a conscious act⁵²³; and (CPR, B 158-9) he suggests that in being "conscious...of [my] power of combination", *what I am conscious of* is the fact that "in respect of the manifold which [I] ought to combine [I] am subjected to a limiting condition (entitled inner sense)" – this limiting condition being the schema. So the conscious awareness of figurative synthesis, as an act *I ought (yet) to perform*, precedes the actual performance of that synthesis and makes us aware of the schema as a constraint on acts of combination.

On the basis of these considerations, we can understand Kant's idea that the (pure) imagination produces the categorial schemata without supposing that the schemata are themselves produced on the basis of a kind of synthesis that occurs (to use Robert Paul Wolff's phrase) in some pre-conscious 'transcendental limbo'. This would absolve us of the task of trying to tell some elusive story about the character of the processes that constitute such synthesis.⁵²⁴ We can also note here, again, that Kant's use of the first-person perspective, and his emphasis on the conscious awareness of objective synthesis as something-to-be-performed (or of a manifold 'I ought to combine'), provide support for my overall interpretive suggestion that Kant considers the conscious perspective of the thinker who is engaged in synthesis essential to his project. These textual data are hard to accommodate if we suppose that Kant offers his account of synthesis from a detached, third-person perspective that aims at theorizing over a set of processes that go on at some sub-personal, pre-conscious level which is beyond our conscious grasp.

I want to conclude this section by relating my disagreement with Longuenesse, concerning the role of the schemata in the objective synthesis of intuitions, to an extremely important idea that she has introduced into Kant scholarship: namely, the notion of a 'twofold application' of the categories. She holds that the categories are first applied 'blindly', *qua* schemata in the synthesis that relates intuitions to objects, and then once more in conscious reflection on experience, where they emerge as 'clearly reflected concepts'. I find the idea of a twofold application of the categories very illuminating, but I think that the triad of categories-schemata-Principles is inseparable: *any* objective employment of the categories must be governed by the Principles as rules and restricted by the schemata. In their first employment in synthesis of given intuitions, the categories make experience in general possible by relating sensible intuitions to empirical objects in general; this allows for the representation of empirical substances and their changes. At a second level, we apply the categories again, and in a more determinate way, when (and if) we get into the epistemically lucky position to cognize *specific kinds of* substances or *specific kinds of* laws of change; this is the level at which the formation and application of empirical concepts occurs. This picture of a twofold application of the

⁵²³ See B 154: "...[time as a unified, determinate intuition] is possible only through the consciousness of the determination of the manifold by the transcendental act of imagination..." – again (cf. section 3 above), consciousness of the synthesis is a *necessary condition* of unity, namely, the unity of time as an intuition.

⁵²⁴ Longuenesse herself concedes that the claim that the mind first produces the categorial schemata independently of a conscious awareness of the categories and principles implies that there is a gap in Kant's discussion in the *Analytic*. She seeks to fill this gap by providing, on his behalf, "a case-by-case explanation of the productive syntheses of imagination as they relate to logicodiscursive forms for which they are produced, thus generating the schemata of pure concepts of the understanding" (2001, p. 245). But I do not really see how Kant could have any *genetic, explanatory* story to tell about the pre-conscious genetics of the noumenal mind (Wolff's "transcendental limbo"); after all, he says (CPR, A 141/B 180-1) that the schematism is "an art concealed in the depths of the human soul, whose real modes of activity nature is hardly likely ever to allow us to discover".

categories is compatible with the view that the 'first' application of the categories to intuitions in objective synthesis *already* involves an awareness of the categories, the Principles, and the schemata. Our application of the concept of causality cannot be dependent on the cognition of specific causal laws, because our awareness of this concept, or of the corresponding rule ('everything that happens has a cause'), is necessary for experience; by contrast, what empirical laws experience allows us to cognize and to systematically correlate is a contingent matter.⁵²⁵

5: Synthesis of perceptions as a free, norm-governed activity

I have tried to show that the Analogies are the rules that govern the (conscious) synthesis of perceptions that brings them to synthetic temporal unity. In this section I want to argue, first, that the Analogies have a normative status; and secondly, that we have a reflective freedom *not* to combine our perceptions in accordance with the Analogies.

When I say that the Analogies have a 'normative status', I mean that they are rules that state how we ought to combine given intuitions, rather than rules that describe how we actually combine given intuitions. This understanding of the status of the Analogies, as rules for synthesis, is implicated by what Kant says at CPR, B 219 (my emphasis):

But since experience is a knowledge of objects through perceptions, the relation [involved] in the existence of the manifold **ought to be represented** [Kemp Smith: "has to be represented"; the German is: "wie es objektiv in der Zeit ist, in ihr vorgestellt werden soll"] in experience, not as it comes to be constructed in time but as it exists objectively in time.

Here, Kant contrasts the notion of 'how the relation in the existence of the manifold objectively is in time and how it *ought* to be represented' with the notion of 'how this relation *comes to be*

⁵²⁵ Longuenesse's claim that the "reflection [or] subsumption [of appearances] under categories as concepts", which (on her account) occurs in the formulation of the principles, "can occur only when empirical judgments...have been formed and systematically correlated" (2001, p. 244) faces a striking interpretive problem: it clashes with Kant's idea, which is very prominent in the Introductions to the Third Critique, that the a priori certainty of the notion that appearances can be subsumed under the principles does not entitle one to infer that it is *also* possible to systematically cognize appearances according to interconnected empirical laws (EEUK, 20: 209; KU, 5: 179-80). I will discuss this point in more detail in chapter XI, section 2. Here I want to note that what Kant says about the 'clear reflection' of the concept of causality is compatible with my reading. Commenting on the consequences of the view that the concept of causality is formed inductively, Kant says (CPR, A 195-6/B 240-1): "the concept, if thus formed, would be merely empirical, and the rule which it supplies, that everything which happens has a cause, would be as contingent as the experience upon which it is based. (...) Certainly, the logical clearness of this representation...is possible only after we have employed it in experience. Nevertheless, recognition of the rule, as a condition of the synthetic unity of appearances in time, has been the ground of experience itself, and has therefore preceded it a priori." This passage leaves no doubt that a, *however basic or dim*, recognition of (Rücksicht auf) the concept of causality and the corresponding "rule which it supplies" (the Second Analogy) is presupposed by the synthesis that brings given intuitions to synthetic unity. Kant's point that 'logical clarity' of the concept of causality requires experience does not imply that the synthesis of intuitions that achieves 'synthetic unity of appearances in time' is not governed by an awareness of the concept of causality (and of the Second Analogy) *at all*. Now, an awareness of a concept must pertain to *some* of its prominent marks (otherwise there would be no object at all for our 'recognition' to attend to); in the present case, the awareness of the concept of causality that governs synthesis must include at least a, maybe obscure, awareness of the marks of 'hypothetical necessity' and 'universality'. But this leaves much to be established for an 'analytic reflection' on the content of this concept that takes place *posterior* to experience and that requires an appeal to examples and observation. This is so because one can always achieve *greater* logical clarity about the content of the categories: analyzing these concepts is an ongoing task that can *never* be deemed complete (CPR, A 729-30/B 757-8).

constructed in time'. With the latter phrase, he is referring to the psychological processes that constitute the 'automatic' course of perceptual events and by virtue of which a merely subjective temporal order is constructed. With the former phrase, he refers to a normative condition on the representation of an objective temporal order. It is very important that Kant holds that the psychological processes he refers to in the latter phrase themselves *occur in time*: this suggests that as empirical processes, they are potential objects of cognition and as such have to stand in determinate temporal relations to one another. *Any* scientific attempt to cognize our mental activity (e.g., our combinatory psychological processes) *presupposes* cognitive norms that prescribe how one ought to combine one's representations *of* these mental acts for the purposes of assigning them a position in an objective temporal framework.

These cognitive norms are the Analogies of Experience: "An Analogy of Experience is, therefore, only a rule according to which a unity of experience ought to arise ["entspringen soll"] from perception" (CPR, A 180/B 223). Kant goes on to say that these rules are regulative "of the objects, that is, of the appearances". These striking claims make sense within the interpretive framework I proposed in the preceding section. I argued that the Analogies are rules that guide the transition from a subjective, empirical unity of conscious perceptual states to an objective, intersubjective unity of experience; likewise, given Kant's Copernican Revolution and its idea that principles that elucidate necessary conditions for the experience of objects are 'ontologically transparent' to the way objects must be, they impose conditions on the objects of experience *qua* appearances (empirical objects *in general, überhaupt*).⁵²⁶ I now want to add to my interpretation the claim that Kant explicitly understands the point of the Analogies in normative terms: they are rules that specify how we *ought to* combine our perceptual representations.

This interpretation can be strengthened by what Kant says, specifically, about the Second Analogy (CPR, A 191/B 236): "in spite also of the fact that their representation in apprehension is always successive, **I ought to show** what sort of a connection in time belongs to the manifold in the appearances themselves" (my emphasis). The Second Analogy is a rule that prescribes *how we ought to* connect up perceptions that always have a successive course in such a way that objective successions can be represented. Now, this statement brings us back to the important textual datum that Kant ubiquitously uses the first-person present tense in the Analytic. In this particular case, one might hold that his use of normative language ('I ought to show') is inconclusive concerning the question of whether the Second Analogy has a normative status, because in saying 'I ought to show...', Kant might be speaking from the perspective of a theorist who aims at indicating, 'from the outside', the rules that govern synthesis 'from the inside'. But I think that the context of the passage undeniably shows that Kant is once again simulating, for the theoretical purposes of justifying rules for combining perceptions, the 'engaged' perspective of the subject of perception who stands in need of governance by rules of synthesis:

I perceive that appearances follow one another, that is, that there is a state of things at one time the opposite of which was in the preceding time. Thus **I am really connecting** two perceptions in time. (CPR, B 233; my emphasis) ...in spite also of the fact that their representation in apprehension is always successive, **I ought to show** what sort of a connection in time belongs to the manifold in the appearances themselves. For instance, the apprehension of the manifold in the appearance of a house **which stands before me** is successive. The question then arises, whether the manifold of the house is also in itself successive. (CPR, A 191/B 236; my emphasis)

⁵²⁶ I will discuss Kant's contrast between 'regulative' and 'constitutive' rules in the next chapter.

Surely, Kant the philosopher can say these things without himself currently having a house before his eyes. Again, pace Kitcher, I do not find anything 'unwitting' in Kant's methodology here. This is because, unlike Kitcher, I believe that synthesis aimed at temporal unity is a conscious mental activity, and because I believe that rules that govern such activity have the status of normative prescriptions. If so, it seems altogether witting or fitting that Kant the theorizer does, in the context of examining the credentials of those rules, simulate the *norm-governed self-awareness* of the subject of synthesis.

At this point it might be asked why one couldn't read Kant as holding that the Second Analogy specifies *both* how we must think as a matter of empirical (psychological) necessity *and* how we ought to think as a matter of rational necessity. That is, one could concede that the rules of synthesis for Kant have a normative status without giving up on the idea that when Kant puts forward these rules he seeks to describe the actual psychological processes of the synthesizing mind. My first response here is rather simplistic: while the normative reading is explicit at the textual level, there is no textual support for the claim that Kant is *also* concerned with describing actual processes by appeal to psychological necessity. But I want to provide a further reason against the claim that Kant seeks to show that the Second Analogy governs how we must (empirically or psychologically speaking) think.

Here we need to take a careful look at Kant's employment of the notion of a rule:

...that which follows or happens must follow in conformity with a universal rule [1] upon that which was contained in the preceding state. (...) That something happens is, therefore, a perception which belongs to a possible experience. This experience becomes actual when I regard the appearance as determined in its position in time, and therefore as an object that can always be found in the connection of perceptions in accordance with a rule [2]. This rule [2], to determine something according to succession of time [etwas der Zeitfolge nach zu bestimmen] is, that the condition [= rule 1] under which an event invariably and necessarily follows is to be found in what precedes the event. (CPR, A 200-1/B 245-6; square brackets are my addition.)

Rule [1] is an empirical rule or condition that governs natural alterations. According to the naturalistic reading, rules of synthesis causally necessitate our mental acts and (hence) rule [2] must be of the same generic type as rule [1]. So, on this reading, Kant is saying that it is an invariable empirical law [2] that subjects react to their perception of events by positing invariable empirical laws that necessitate the perceived event.⁵²⁷ There are various reasons to think that this misconstrues Kant's understanding of rule [2]. First, Kant refers to rule [2] as a rule for us *to determine* some event according to the objective succession of time rather than as a rule by which we do in fact determine the temporal order. Second, Kant appeals to the kind of necessity and invariability that is distinctive of empirical laws *only* by reference to rule [1], *not* by reference to rule [2]. Third, the naturalistic interpretation that assimilates rule [2] to rule [1] gives rise to a vicious circle. Kant clearly attempts to give the notion empirical laws, and thereby the 'must' of empirical necessity, a grounding by appealing to the way in which we *must* (in some sense of 'must') react to the perception of events.⁵²⁸ But according to the naturalistic reading, our

⁵²⁷ More precisely and tortuously: there is an empirical law [2] that invariably and necessarily governs a mental event [a] following from a preceding event [b]; the mental event [a] being governed by empirical law [2] is the event of positing an empirical law [1] that invariably and necessarily governs the occurrence of a perceived event [c] following from a preceding event [d].

⁵²⁸ This is accepted by the proponent of the psychological reading. See Kitcher 1999, p. 434: "...the causal relation is itself contributed by the actions of our faculties".

reaction to perceiving events is itself governed by an empirical law, and our synthesis of perceptions is itself subject to empirical necessity. Hence, on this reading, Kant attempts to explain the subjective origin of empirical law and necessity by appealing to the way in which we must, as a matter of empirical law and necessity, synthesize perceptions. And this seems circular.⁵²⁹ One cannot have it both ways, that is, both insist in an idealist fashion on the subject-dependency of the causal relation and pretend, in a realist-naturalist fashion, that the processes on which the causal relation allegedly depends can themselves be characterized *as causal processes* and hence as already exhibiting the very property they are meant to 'contribute'. If the modalities employed in the rule that says that we must combine perceptions in a certain way are meant to be of a piece with those represented in causal laws of nature, then that rule cannot purport to ground or explain the 'must' of causal necessity. Hence, I think that we should not read Kant as offering his rules of objective synthesis – such as the Second Analogy [rule 2] – as psychological laws that represent the 'must' of causal, empirical necessity.⁵³⁰

I have argued that the Second Analogy is a prescriptive norm addressed to the human subject of synthesis, informing her that the representation of an objective succession (an alteration) requires positing a causal ground that brings about this succession in accordance with an invariable, necessitating law. I now want to address a very important worry about the idea that Kant's rules for synthesis are intended as prescriptive norms in the manner envisaged by my interpretation. This worry has been aptly put to me by Beatrice Longuenesse as follows: there is no need for Kant to *establish* norms that govern our thinking in terms of causal necessity, *because we think in these terms anyway*. Newton did not (have to) wait for Kant to come along to suppose that events are related through necessitating causal laws.

It seems fair to say that human beings have exhibited *a tendency* to attribute observed

⁵²⁹ Kitcher is aware of the problem (see her 1999, p. 434 ff.), but she thinks that the only thing that this "reapplication of the model to itself" shows is that "Kant's epistemological theory has...the same status as a scientific theory". For an ordinary scientific theory that simply takes the concept of causal, empirical necessity for granted and looks for laws that instantiate it, there is indeed no problem here. But a scientific theory that aspired to uncover the psychological ground of the concept of causal law and necessity could not simply avail itself of that very concept. Imagine, by analogy, a chemical theory trying to uncover the mechanisms underlying evaporation by appealing to the mechanism of evaporation. In both cases the 'explanation' invokes the very property it is supposed to account for. If the idea that "the causal relation is itself contributed by the actions of our faculties" is supposed to have the status of a causal, explanatory claim, it is viciously circular or (what amounts to the same) explanatorily empty.

⁵³⁰ The need to distinguish between various senses of 'rule' in CPR, A 200-1/B 245-6 is also seen by Longuenesse 2001, pp. 368-71. My understanding of rule [1] is identical to here rule 'r' (designating an empirical, causal law). However, Longuenesse posits *three* rather than two meanings of 'rule'; she holds that the rule that governs our representation of an alteration is not the Second Analogy itself, but rather the schema of causality as a rule of sensible synthesis (her 'r*'). As I tried to show, the Second Analogy itself is the rule that governs our synthesis of perceptions that allows for the representation of an objective succession. Longuenesse claims that the Second Analogy (her rule 'R') is to be understood as follows: "The statement that 'everything that happens presupposes something else upon which it follows according to a rule' does not mean that we cognize this rule, but that we are so constituted as to search for it" (1998, p. 366). I fully agree that complying with the Second Analogy is prior to the cognition of the empirical law that one posits in virtue of such compliance, but I do not think that the Second Analogy makes a descriptive claim about our psychological dispositions. Elsewhere (2009, pp. 168-9), Longuenesse proposes a normative understanding of the Second Analogy; she says that it is a principle "that says that we should look for rules to confirm our perception of something haven happened". But I do not think that the Second Analogy quite says that. What it says is that we ought to *posit* ("setzen"), for any perceived alteration, an invariable causal law.

alterations to the workings of empirical causes that are governed by natural laws. But what are the grounds of this tendency? Here are two possibilities: One may think that events are necessarily connected (1) on the basis of a (maybe implicit) conviction that there must be a sufficient reason for anything that happens in nature; or, as Kitcher and Longuenesse suggest, (2) on the basis of blind mental processes that occur regardless of any such conviction.

Now, notice that (2) is very reminiscent of *Hume's* account of the 'basis' on which we represent events as causally connected; and hence Kitcher and Longuenesse seem to be construing Kant as a 'quasi-Humean'. For sure, on their view Kant's theoretical ('outside') perspective on the subject of synthesis is more conciliatory than Hume's: Kant attributes the concept of causality to *productive* rather than reproductive imagination, and he considers it a valid concept. But, on this interpretation, Kant agrees with Hume about one important point: namely, like Hume he thinks that our experience of events as causally connected is removed from our awareness of any reasons for regarding events as so connected. This implies that Kant shares Hume's view that what leads us to use concepts that are crucial for the way we represent the world are our given mental dispositions that (almost literally) force us to think in the way we do, and that we are, *in this respect*, quite like animals.⁵³¹ I think that attributing such quasi-Humeanism to Kant accords neither with his general picture of what is distinctive of the *human* (as opposed to animal) mind nor with his specific methodology in the *Analytic*, which, as we saw, involves a constant appeal to the conscious perspective of the synthesizing subject.

A less 'animalistic' (or mechanistic) construal of our tendency to attribute natural events to causal, lawlike regularities would allow that this attribution is reflective (if only implicitly so, in the sense that our reasons for making this attribution operate 'dimly' in the background of thought) rather than blind. Here we can notice what is arguably a gap in *Hume's* argument for the view (2) that the processes that lead us to use the concept of cause are 'blind': he appears to infer (2) from his conviction that there is no *good* (non-circular) argument that supports the idea that events are necessarily connected.⁵³² One might object that he should at least have taken into account the possibility that we apply the concept of cause on the basis of *bad* reasons that we misconceive as good ones. This brings us to (1): it seems not at all implausible to conjecture that if someone responds to her perception of an alteration by positing some necessitating causal ground, this is based on some (maybe implicit or obscure) conviction that reality as such is fundamentally intelligible, that there must be a sufficient reason for *anything that happens*. Now, if this conviction informs and motivates our synthesizing in causal terms, then our synthesis is, in Kant's view, potentially *misinformed*: for this conviction might betray a commitment to the dogmatic ('realist') assertion that anything that could possibly exist or happen must be governed by natural necessity. This might well be Kant's verdict about Newton: Newton's positing of

⁵³¹ Kitcher and Longuenesse would hold that animals have neither spontaneity nor productive imagination. But in severing the tie between spontaneity and conscious awareness, they seem to me to narrow the gap between the human and the animal mind: for surely Kant thinks that it is, above all, the capacity for (*self-*)*conscious*, (*self-*)*reflective* mental activity that is distinctive of the human mind. It is not clear what difference it makes to call an exercise of imagination productive or reproductive if, for the subject in whom the respective workings of the imagination occur, there is no difference that is accessible to consciousness. I return to this point below.

⁵³² Millican 2002, p. 132 notices that when Hume says that the causal inferences are not based on reasons, "the qualification 'good' is usually implied". Now, this is no problem if Hume wants to establish only that the fiction-inducing imagination is involved in causal inference; but if he seeks to show that we make these inferences altogether blindly or unreflectively, his argument is inconclusive in that it leaves out the possibility that our causal inferences are formed reflectively on the basis of an awareness of reasons we falsely represent as good.

causal connections is rationally deficient because it is based on mistaken, realist grounds.⁵³³

These considerations show that *even if* we assume that humans synthesize in terms of the concept of causality *anyway*, it is not implausible to read Kant as offering his rules of synthesis as norms that are meant to influence reflective acts of synthesis. The idea that the act of synthesizing perceptions in terms of the concept of causality has a reflective dimension prevents Kant's account from collapsing into the kind of quasi-Humeanism that is inherent in (2). The idea that Kant has the ambition to elaborate a norm for synthesizing in causal terms can be motivated by appeal to the fact that he rejects the realist convictions that tend to dominate people's acceptance of the notion that events are causally connected. In particular, the norm he offers puts explicit emphasis on the fact that the justificatory basis for referring successions to necessitating grounds derives from the fact that an objective temporal order is (contrary to what the realist picture assumes) something that has to be constructed by the thinking subject in the first place. Thus, while Longuenesse is right to emphasize that Kant does not take himself to invent anything that is not already contained (however obscurely or implicit) in the structure of any human understanding, I would stress that he does take himself to make explicit, in systematic terms, the *proper* justificatory basis for the demands of the understanding, a basis which he deems accessible to the synthesizing subject.⁵³⁴

But now I want to suggest that reading Kant as offering his rules of synthesis as norms addressed to synthesizing subjects is recommended primarily by the fact that he does *not* believe that human subjects experience events as causally connected anyway: he holds that human beings, at least when considered from what I have called the normative standpoint (cf. chapter IX, sections 3-4), have a kind of freedom *not to* think of events as causally connected. The Second Analogy, qua norm of synthesis, is meant to counteract the tendency underlying this freedom. Here we must take into account what Kant says at CPR, A 196-7/B 241-2:

We have, then, to show, in the case under consideration, that we never ourselves in experience ascribe succession (that is, the happening of some event which previously did not exist) to the object, and so distinguish it from subjective sequence in our apprehension, except when there is an underlying rule

⁵³³ Of course, Kant's disagreement with Newton *becomes apparent* only at a level whose cognitive determinacy and conceptual sophistication goes much beyond the level of original synthesis at which we merely relate intuitions to empirical objects in general: namely, at the level where he transforms Newton's realist framework in which absolute space and motion are being presupposed into what Michael Friedman aptly calls "a constructive procedure for first defining the concept of true motion" (1992b, p. 143). But this is no objection to the idea that Newton's realism has a bearing already on his synthesizing of intuitions, i.e., on his practice of referring of perceived events to necessitating grounds. The original synthesis of given intuitions is, after all, a fundamental cognitive starting point that naturally bears on our subsequent, conceptually more determinate cognitive activity. If one relates perceived events to necessitating causal grounds on the basis of mistaken realist premises, a seed has been sown for disastrous results that shall emerge only at the level of conceptually determinate cognitive activity: namely, when the positing of space and time as things in themselves leads reason into internal conflicts (antinomies). See my chapter IX, section 4, and chapter XI, section 3.

⁵³⁴ I want to say something analogous about the practical case. Longuenesse has pointed out to me that attributing to Kant the intention to give prescriptive *moral* advice is problematic because Kant did not take himself to 'invent' new moral ideas. Again, I think that she is right to emphasize this latter point, but it is equally important to see that Kant saw a dire need to articulate clearly the ideas that are only obscurely contained in common moral sense, so that they can function as proper standards of practical justification and be protected from infusion with sentimentalism or from misinterpretation (cf. the standard identification of the formula of the universal law with the Golden Rule). He thereby meant to introduce a procedure for testing maxims that is *novel at least in its systematicity*, and by that he clearly meant to influence volition.

which **compels** us to observe this order of perceptions rather than any other; nay, that this **compulsion** is really what first makes possible the representation of a succession in the object. (My emphasis.)

Kemp Smith's translation here is misleading, for the term 'compulsion' might be taken to indicate a kind of natural unavoidability that parallels the sense in which a body cannot avoid falling when dropped. This brings to mind not only Kitcher's appeal to the law of gravity but also Paul Wolff's idea that the objective synthesis of perceptions is "pre-conscious as well as unavoidable".⁵³⁵ But it is important to note that the word that Smith translates as 'compulsion' is 'Nötigung', and in Kant this term is ambiguous (an ambiguity corresponding to the ambiguity in his use of the term 'rule' at CPR, A 200-1/B 245-6, as discussed earlier). When Kant talks about Nötigung (compulsion or necessitation) through rules, he *may* refer to the necessitation of an empirical object by causes that are governed by nomological regularities. But he may also refer to the relation between an imperfectly rational subject and an objective normative law of reason (GMS, 4: 413): "All imperatives are expressed by the word ought, and thereby indicate the relation of an objective law of reason to a will, which from its subjective constitution is not necessarily determined by it (eine Nötigung)." In other words, the term 'Nötigung' signals, in connection with normative rules that express an 'ought', an action that is only *rationally unavoidable*, but that imperfectly rational thinkers are free to avoid. The Second Analogy *is* a rule, directed to the understanding (not to the will!), that expresses an 'ought' ("a rule according to which unity of experience ought to arise from perceptions", CPR, A 180/B 223). Hence, it indicates the relation of an objective law to an understanding which from its subjective constitution is not necessarily determined by it. Hence, we – considered as spontaneous thinkers – are not causally necessitated to posit necessitating causes when we perceive events. The term 'compulsion' here signals a rational unavoidability rather than material necessity, just as the term 'rule', as applied to the causal principle, refers to a normative rather than to a causal law.

Notice that this does not imply that Kant is an epistemic voluntarist. As we saw in chapter IX, Kant rejects the idea that we can judge or think contrary to our conscious recognition of how we ought to think or judge. So our freedom not to comply with rules of synthesis cannot amount to a voluntary control over our synthesis that would allow us to *choose* to perform acts of synthesis contrary to our awareness of how we ought to synthesize. Instead (cf. chapter IX, section 4), this freedom relates to our tendency to come under the influence of factors that corrupt our conception of how we ought to think or synthesize and thereby corrupt our acts of thought of synthesis. I now want to spell out what such corruption amounts to.⁵³⁶

⁵³⁵ Wolff 1963, p. 131 (cf. footnote 511 above); Henrich's 1976 conception of acts of synthesis is similar to Wolff's in that Henrich also thinks that we cannot help but complying with rules of synthesis.

⁵³⁶ According to Anderson, 2001 p. 292, our representations "fail to match the general norms of synthesis" when "we dream, have perceptual illusions, make false judgments, etc." I think we need to be careful here. *All* of our passively received representations *as such* fall short of constituting objective experience; this point applies to the representations induced by a dream just as it applies to the mere perception (in Kant's technical sense) of a ship moving downward. These representations 'fail to match' norms of synthesis simply because they have not yet been combined by the act of spontaneous synthesis that is aimed at experience of empirical objects in general and that is governed by these norms (we can add that the representations induced by a dream are intrinsically unsuitable for such combination). When we ask for an example in which *an act of combination is governed by and violates* the norms of synthesis, I do not think that either of the three cases that Anderson mentions qualifies. First, the combination of representations that occurs when we dream is, rationally speaking, *too thin* to qualify as standing under the governance of normative rules. Likewise (second), having an illusory experience, such as representing a straight stick in the water as bent, is not governed by a priori norms of synthesis either, for the

The rule for determining the sequence of time says that an alteration "can acquire [a] determinate position in this [time] relation of time only in so far as something is presupposed in the preceding state upon which it follows invariably" (CPR, A 198/B 243). This implies that one *fails* to represent an objective succession if one *fails* to presuppose or posit a ground from which the alteration, invariably and necessarily, follows.⁵³⁷ This can happen under the influence of one of the factors that are (cf. chapter IX, section 4) chiefly responsible for the imperfection in our epistemic rationality: namely, the factor of transcendental illusion, whose ever present influence (arising from our tendency to mistake appearances for thing in themselves) may affect even the kind of synthesis that epistemically precedes conceptually determinate empirical judgment.

I want to indicate two ways in which the influence of transcendental illusion might lead astray (via a corruption of our conception of how we ought to synthesize) our attempts to transform our merely subjective, empirical unity of consciousness into objective experience. First, and very prominently, the rule for representing objective succession will be violated by a thinker who posits an unconditioned causality as the real, explanatory ground of events within the natural order: this positing may rest on the mistaken belief that all empirical conditions must, on pain of a vicious explanatory regress, be grounded in something unconditioned (this is the thesis of the Third Antinomy); or it may rest on the confused notion that we can find a place for our absolute volitional spontaneity, as a ground of free actions, within the natural order of things, i.e., the notion that our practical freedom is an object of sensible intuition that we can appeal to for the purposes of explaining why we act as we do (this is the tendency to 'sensualize', the idea of spontaneity that Kant laments at GMS, 4: 452). Here an alteration is referred to a condition that does not operate according to an empirical law and hence does not prompt its effect invariably and necessarily. It is important to realize that there is conclusive textual evidence for the claim that Kant holds that the influence of the transcendental illusion that leads one to posit spontaneity within nature affects not only one's high-level, scientific theorizing about nature but also one's cognitively basic attempts to get a handle on natural objects *in general*. In the *Analytic*, Kant describes the consequences of *not* positing ("setzen") a ground of succession that causes its effect invariably and necessarily as follows:

I should have to regard the succession as a merely subjective play of my fancy; and if I still represented it

subject of synthesis cannot help but seeing the stick as bent even if she is aware of the illusion; the illusory representation merely happens to her. By contrast, third, making a false empirical judgment such as 'the stick is straight' is *too thick* to qualify as an incorrect act of the kind of synthesis that relates given intuitions to objects. For making such a judgment is an instance of an (mistaken) ascription of an empirical property to a substance, and this act of 'determinate thought' goes beyond the cognitively basic acts of thought governed by rules for the kind of synthesis that precedes all analysis. In ascribing the property of being bent to the stick, we successfully draw the distinction between something that underlies changes (the stick) and something that goes away in changes (being bent) – hence, we comply with the First Analogy, succeed in representing an intersubjectively accessible object, and thereby get into a position to make true or false judgments of experience in the first place.

⁵³⁷ I want to be clear that my question here is *not* whether we have the reflective freedom not to think of two types of events that we find to follow one another repeatedly as causally connected. I think it is obvious that we have such freedom; even Hume must admit this, for he must allow that we can distinguish between a mere correlation and a causal connection (whether he can make sense of that distinction is a different matter). Determining whether two regularly conjoined types of events are merely correlated or causally connected is not the point of the 'original' synthesis that determines an alteration as an object of experience; rather, this issue arises only at the level of conceptually determinate systematic scientific theorizing. My question is, thus, whether we are free not to posit *some* empirical cause and causal law, whatever it is, for a perceived alteration.

to myself as something objective, **I should have to call it a mere dream.** (CPR, A 201-2/B 247; compare CPR, A 194/B 239; my emphasis)

In the Dialectic, Kant describes the consequences of the transcendental illusion that leads one to credit natural objects with spontaneity as follows:

But to ascribe to substances in the world itself such a power [a transcendental power of freedom], can never be permissible; for, should this be done, that connection of appearances determining one another with necessity according to universal laws, which we entitle nature, and with it the criterion of empirical truth, **whereby experience is distinguished from dreaming,** would almost entirely disappear. (CPR, A 451/B 479; my emphasis)

These passages show that for Kant the influence of transcendental illusion extends to the synthesis which, if performed properly, allows us to represent an objective temporal and natural order; the influence of transcendental illusion that leads us to ascribe spontaneity to empirical substances prevents us from achieving this goal. While someone who has succumbed to this influence may still think that she has succeeded in representing an objective temporal order, her mental state really is something akin to 'dreaming' or 'the subjective play of her fancy'.

In addition to this *rationalistic* corruption of our conception of how we ought to synthesize, we can mention a second *empiricist* corruption of this conception. The rule that prescribes that one must refer an objective succession to an invariable and necessitating causality will *also* be violated by someone, like Hume, who fails to see how we could know that there might be real necessary connections between events, and who tries to debunk our positing of such connections by appeal to subjective habits of the imagination.

These are only two ways in which transcendental illusion accounts for a possibility that we might fail to comply with objective rules of synthesis; since susceptibility to transcendental illusion is 'written into' our imperfect cognitive faculties, it is only fitting that Kant designates the Second Analogy, qua rule of synthesis that represents an objective demand of empirical thought to a thinker who from her subjective constitution is not necessarily determined by it, as "a rule according to which a unity of experience *ought* to arise from perception".⁵³⁸

Apart from the freedom to violate normative rules of synthesis that we possess in virtue of our rational imperfection, it should be noted that we can quite *legitimately* choose to set aside these rules and (hence) deliberately opt out of the activity of experiencing an objective world. It is perfectly legitimate, even rationally unavoidable for us to leave behind (on occasion) the empirical standpoint of objective experience and instead adopt the normative standpoint of freedom, from which we attribute (*albeit not for theoretical, e.g., explanatory purposes*) our phenomenal acts to a transcendental cause that does not operate according to empirical laws.

I want to resume the discussion in this section by looking at a very important passage from the Deduction where Kant reflects on the possible origin of the categories:

A middle course may be proposed between the two above mentioned, namely, that the categories are ...subjective dispositions of thought, implanted in us from the first moment of our existence, and so ordered by our Creator that their employment is in complete harmony with the laws of nature in accordance with which experience proceeds -- a kind of preformation-system of pure reason. (...) there is this decisive objection against the suggested middle course, that the necessity of the categories, which belongs to their very conception, would then have to be sacrificed. The concept of cause, for instance,

⁵³⁸ A commitment to transcendental realism might lead one either to comply with the rules for synthesis on mistaken grounds (Newton's case) or to violate these rules. Transcendental realism is a hydra with many heads.

which expresses the necessity of an event under a presupposed condition, would be false if it rested only on an arbitrary subjective necessity, implanted in us, of connecting certain empirical representations according to the rule of causal relation. I would not then be able to say that the effect is connected with the cause in the object, that is to say, necessarily, but only that I am so constituted that I cannot think this representation otherwise than as thus connected. This is exactly what the sceptic most desires. For if this be the situation, all our insight, resting on the supposed objective validity of our judgments, is nothing but sheer illusion; nor would there be wanting people who would refuse to admit this subjective necessity, a necessity which can only be felt. Certainly a man cannot dispute with anyone regarding that which depends merely on the mode in which he is himself organised. (CPR, B 167-8)

I think that this passage has a crucial bearing on the question of whether Kant conceives of our synthesizing in causal terms as a blind, (quasi-) mechanistic process.⁵³⁹ For I worry that this mechanistic interpretation of rule-governed synthesis saddles Kant with the kind of picture that he rejects toward the end of the above passage. The notion that 'I am so constituted that I cannot think this representation otherwise than as thus connected' *is*, according to Longuenesse and Kitcher, characteristic of any human thinker. Longuenesse might respond that this notion is problematic only under the assumption that the necessity to think in (e.g.) causal terms is implanted in us from the outside. But it is clear that Kant's worry has a broader scope: the notion of 'subjective necessity' that he uses in the above passage standardly relates to the Humean idea that we are psychologically compelled to think in causal terms 'from the inside'.⁵⁴⁰ And my worry is that on Longuenesse's and Kitcher's reading, Kant's picture is, from the perspective of the synthesizing subject, *indistinguishable* from the picture offered by the Preformation Theorist or by Hume. The common core of all three pictures is that synthesizing in causal terms is *something that happens to us* at a sub-personal level, that is removed from our reflective control, and that we cannot help but do. The notions that are distinctive of mental activity that has a rational dimension – such as: wondering, doubting, misconceiving – are here rendered inapplicable to our practice of representing events as necessarily connected. What difference, then, does it make whether, according to an external diagnosis, these mental happenings are caused (as on Longuenesse's and Kitcher's picture) by the productive imagination rather than (as in Preformation Theory) by an external agency or (as in Hume) by the reproductive imagination?

Here is another way of making this point. According to Kitcher's and Longuenesse's view, the normative dimension of our synthesizing under the concept of causality derives from the fact that the theorist (i.e., Kant) can diagnose, from the outside, that thinking in causal terms

⁵³⁹ I think that Guyer 1987, pp. 368-9 misunderstands the significance of this passage. He equates Kant's distinction between the idea that the categories have their origin in the spontaneity of the mind and the idea that the categories are subjective necessities implanted in us with his distinction between (what he calls) 'absolute' (imposed) and 'conditional' necessities. A conditional necessity, for Guyer, is of roughly the form, 'If we are to represent an object, we must represent it as satisfying condition C', where this cannot (on Guyer's view) establish *de re* necessities about the way objects must be (since, on Guyer's view, Kant has no good argument for transcendental idealism). Guyer claims (*ibid.*, p. 349) that such a conditional necessity does provide evidence for the way objects actually (if only contingently) are, but this is a moot point: the hypothesis that we must think about objects in a certain way leaves open whether there is a cognitive fit between these ways of thinking and the actual constitution of objects; Kant would surely insist that within a realist framework this gap cannot be bridged. This worry becomes even more pressing given Guyer's contention that Kant has no way of ruling out that the categories are implanted subjective necessities: for this implies that the way we must think about objects may just an arbitrary fact about the way we happen to be constituted, quite regardless of the actual constitution of objects. And this deprives the conditions of thought that Kant seeks to establish of all cognitive significance.

⁵⁴⁰ See CPR, B 5, B 167-8, A 94/B 126-8; and KpV, 5: 12-13.

is the correct thing to do: this verdict about the justifiability of thinking in causal terms is irrelevant to the activity of the subject who engages in this kind of thought, since the subject's going on as it does is a blind process that is as insensitive to there being any reasons for going on in this manner as the movements of the planets.⁵⁴¹ But this looks very much like the scenario imagined by the Preformation theorist. For here, too, our synthesizing can be deemed 'correct' or 'appropriate' from the external viewpoint of the Divine Agent who wires us in such a manner that our synthesis "is in complete harmony with the laws of nature" that the Divinity infallibly intuits. Let us imagine that the Divine Agent recognizes that subsuming intuitions under the concept of causality is (for thinkers like us) necessary for cognition, and therefore benevolently sets up our mind in such a way that we cannot help but subsuming intuitions under this concept. It would seem that the external justification of the way we synthesize is exactly the same here as on the picture endorsed by Kitcher and Longuenesse. Moreover, there is a further similarity between their picture (on which we blindly synthesize under the influence of productive imagination) and the scenario in which we are wired to think in causal terms by a Divine engineer: in both cases our synthesizing under the concept of causality is completely independent of any grasp, on *our* part, of any reasons for using this concept. Kant asserts that in the scenario imagined by the Preformation theorist, our use of this concept would not be rationally justified and would lack objective validity. But if this scenario is so much like the picture that results from Kitcher's and Longuenesse's interpretation, it seems that on their picture the categories are not validated either.

I want to bring out this intuition in a final way. Imagine a computer that is programmed to process information so as to indicate the temperature. We can say that the computer functions correctly if it reliably tracks the temperature. So a stance from which one diagnoses the correctness of acts of processing data can be adopted toward objects whose 'spontaneity' is (as Kant puts it) that of a 'turnspit'. Moreover, it seems clear that whatever standard of correctness we can apply, from the outside, to the operations of such turnspits is, in a sense, irrelevant to how they operate. Of course, what the machine does is – if it functions correctly – sensitive to there being a standard of correctness *in the sense that* what the machine does is the result of it being

⁵⁴¹ On Longuenesse's – as opposed to Kitcher's – interpretation, the notion that the processes by means of which we synthesize given intuitions under (e.g.) the concept of causality are blind may be said to be tempered by her frequent suggestion that there is a sense in which said processes are goal-directed. For instance, she intimates (2001, p. 122) that there are "silent judgments" that "preside over" the blind syntheses that bring given intuitions under the categories (or schemata); and this, I take it, is connected to her notion that these processes (ibid, p. 245) "aim at reflecting the sensible given under concepts combined according to the logical forms of judgment" or (p. 246) proceed "with a view to subsuming appearances under concepts". I think that Longuenesse's overall account shows in an admirable fashion that the synthesis which precedes analysis must be directed by the *focal goal* of all activity of the understanding, namely, the (eventual) formation of conceptually determinate judgments about the world. But her appeal to the *intentional or goal-directed* character of sensible synthesis seems to me to smuggle in an element of conscious awareness that such synthesis is lacking on Longuenesse's official story. How can an unconscious act be said to "aim at" reflecting sensible intuitions under concepts? I can think of cases where the application of goal-directed terminology to unconscious processes makes some sense: for instance, we can say that physical processes in our stomach 'aim at' digestion. But this is metaphorical talk, and it gives us the wrong model for an "intellect-driven combination of representations" (p. 71). Correspondingly, I do not know what to make of the metaphorical notion of a 'silent judgment' 'presiding over' a blind, unconscious process. I understand the notion that an *implicit* judgment rationally informs an act of explicit thought, if it could be articulated by the agent of that thought (i.e., could be carved by her from the reflective background of her thinking processes); but this cannot be the model that Longuenesse has in mind, for on her account there is no judgment to be articulated and the relevant thought process is fully 'blind' or pre-conscious.

programmed in a certain way, and the way in which it was programmed is sensitive to the programmer's awareness of a standard for reliable temperature-tracking. The sense in which the standard is *inaccessible* to the machine, or *irrelevant* to what it is doing, is borne out by the fact that we cannot say about the machine that it is influenced to go on as it does by an awareness of the correctness of the standard, or, more generally, by an awareness of any reason to proceed in one way rather than another. But we cannot say this about the blind workings of our productive imagination on Kitcher's and Longuenesse's interpretation either.

Kitcher and Longuenesse might hold that it does not follow from their interpretation that there is no way in which we can go wrong in our synthesizing, and they might suggest that the reason why it makes sense to theorize about our synthesis in terms of 'ought' judgments is that there may be cases in which we do *not* synthesize correctly. But again, the same is true with respect to the computer, which may be said to process information improperly or incorrectly. Because the computer's synthesis has no reflective dimension, its synthesizing improperly is the result of a mere malfunctioning. By analogy, whenever our blind acts of synthesis fail to occur 'as they ought to', this is a mere 'happening' in the space of causes that lacks a rational dimension.

The analogies between the way in which we synthesize on Kitcher's and Longuenesse's interpretation and the way in which a mere turnspit synthesizes suggest that their interpretation faces a *reductio*: an interpretation of Kant that assimilates our spontaneous synthesis to that of a computer is, for this reason, unsatisfactory. Of course, this line of argument depends on the intuition that Kant would not be happy to assimilate our spontaneous synthesizing activity to that of a very sophisticated machine. I think that the overall argument of this chapter does lend some support to this intuition: Kant, in expounding the rules for objective synthesis, constantly appeals to the conscious first-person perspective of the synthesizing subject shows that he thinks that our (spontaneous) synthesis of given 'data' has a dimension that is lacking in the case of computers (or in the case of higher animals). It is in this conscious, reflective dimension that I locate the possibility of and the need for prescriptive norms ('oughts') that govern epistemic activity.

6: The autonomy of the understanding and the more than human validity of the categories

It is a crucial aspect of Kant's epistemology that an (if only basic or dim) awareness of fundamental concepts such as causality and of the corresponding principles ('every event has a cause') is not formed on the basis of experience, because these concepts and principles make experience of objects (i.e., the representation of empirical objects in general) possible in the first place. I have argued that the cognitive activity (synthesis of intuitions) that is governed by these rules is essentially conscious and reflective. This means that our awareness *of the validity* of the notion that every event has a cause is decisive for our subsumption of perceived events under the concept of causality (in contrast to the way in which a computer combines its data). I now I want to characterize the *standpoint* from which the validity and necessity of the pure concepts and principles of the understanding can be established or recognized.

Here I want to begin by noticing that according to Kant, an *enlightened* subject of thought would conceive of herself as the author of principles such as, 'every event has a cause'. That is, she would recognize that it is her own cognitive faculties that are the source of these principles. The notion that every event has a cause is not dictated to us – on the basis of observation or perception – from the outside; nor is our commitment to this notion based on the contingent workings of our psychological constitution. In this sense, the understanding as a

source of prescriptive laws seems to be on par with practical reason as a source of prescriptive laws, which is why Kant considers both the understanding and practical *autonomous*.⁵⁴² Of course, this conception of ourselves as autonomous authors of epistemic principles will be rejected by someone who is still in the grip of the metaphysics and epistemology implied by transcendental realism – this is why I said that an *enlightened* subject of thought would conceive of herself as the autonomous author of the causal principle. The same is true in the practical case: someone who is still in the grip of, say, Divine Command Theory is not going to share the view that the moral standards and principles she applies have their source in practical reason.

There is a crucial difference between the legislation of the understanding and that of practical reason: the former, but not the latter, pertains to the constitution of natural objects. The understanding prescribes laws *to nature* (CPR, A 126-7; B 164-5). Still, and somewhat paradoxically, I think that the prescription of laws to nature can be considered as an act undertaken from the *non-empirical* standpoint. In chapter VIII, I suggested that Kant's non-empirical, normative standpoint is not solely a standpoint from which one recognizes the validity of practical norms (for volition) but also a standpoint from which one recognizes the validity of norms for empirical thought, a recognition that is prior to and independent of what empirical thought itself may establish. In the light of what I argued in this chapter, we can now see that these norms include the principles of the understanding qua rules for the objective employment of the categories in a synthesis of given intuitions that enables us to have objective thoughts about the natural world. Since thinking under the guidance of these norms is a condition for having objective thoughts about nature, these norms cannot be established on the basis of the observation, explanation, or prediction of natural objects and events. The independence of the principles from empirical observation accounts for their status as *laws* that express an a priori necessity.⁵⁴³ The seemingly paradoxical idea that principles that are recognized as valid from the normative standpoint express descriptive claims about what natural objects in general must be like makes sense if one keeps in mind (cf. section 4) that for Kant the *only* way in which we could establish necessary truths about objects is via a reflection on norms for synthesis that provide necessary conditions for the experience of objects. This line of thought requires the 'Copernican Revolution' in metaphysics which makes objects of experience dependent on the conditions for having experience and hence on our cognitive faculties. On this assumption, the understanding is a faculty whose laws (principles) impose demands both on how given representations must be combined and on what objects of experience must be like. If this assumption is rejected, necessary principles about objects are rendered unknowable.⁵⁴⁴

⁵⁴² See EEUK, 20: 225. See also my chapter IX, section 6, and chapter VIII, sections 3 and 4.

⁵⁴³ The contrast here is between necessary epistemic laws and contingent empirical concepts or *rules*; or between epistemic legislation and mere empirical cognition on the basis of inductive experience. See KU, 5: 174: "Empirical concepts have, therefore, their territory, doubtless, in nature as the complex of all sensible objects, but they have no realm (only a dwelling-place, domicilium), for, although they are formed according to law, they are not themselves legislative, but the rules founded on them are empirical and, consequently, contingent."

⁵⁴⁴ See CPR, B 163-4: "Categories are concepts which prescribe laws a priori to appearances, and therefore to nature, the sum of all appearances. (...) That the laws of appearances in nature must agree with the understanding and its a priori form, that is, with its faculty of combining the manifold as such, is no more surprising than that the appearances themselves must agree with the form of a priori sensible intuition. For just as appearances do not exist in themselves but only relatively to the subject in which, so far as it has senses, they inhere, so the laws do not exist in the appearances but only relatively to this same being, so far as it has understanding. Things in themselves would necessarily, apart from any understanding that knows them, conform

The principles of the understanding are the *only* claims about objects of nature whose truth and validity is not established or recognized from the standpoint of empirical observation. The peculiar (idealist) notion of there being truths about the natural world that are established in complete independence of empirical observation is somewhat moderated by the admission that these principles only give us a most abstract insight into the character of natural objects; in order to go beyond this indeterminate picture, and to gain insights about the specific character of natural objects and about the specific laws governing events, we need to resort to experience. I have, in the course of this chapter, repeatedly tried to do justice to this distinction by holding that the laws of synthesis that the understanding prescribes to the synthesizing subject constrain merely the *general* character of appearances, and (hence) that compliance with these laws allows merely for the cognition (experience) of empirical objects *in general*, independently of any further determination of these objects through empirical concepts and laws.⁵⁴⁵ Such further cognition depends on what is given to us through affection by objects, and hence we are not here occupying the normative standpoint of autonomy but the standpoint of empirical observation. I will return to this important contrast in the next chapter.

In chapter VIII, I argued that the non-empirical, normative standpoint is a 'more than human' standpoint, in two respects: first, it involves representations that are shared by all thinking beings; second, it relies on justificatory standards that are binding on all thinking beings. I now want to suggest that this notion of a more than human perspective can be applied to the perspective from which one recognizes the validity of the categories as sources of cognitive norms. First, the validity of these concepts is recognized from a standpoint from which we are representing a discursive intellect that depends for objective thought on a given sensible manifold that must be spontaneously combined in a specific manner. This representation of beings that are sensibly affected but not determined (and hence capable of *spontaneous* combination) is a 'more than human' representation, because it is independent of the forms of human sensibility. It applies to the "other thinking beings" that Kant mentions in the Aesthetic (CPR, A 28/B 44), which may not be bound by "the conditions...that restrict our sensibility".

Second, the justification of the epistemic principles that 'speak to' the task of a discursive understanding is based on standards of thought that are shared among all sensibly affected thinkers: the pure, unschematized categories. The first part of the B-Deduction bears out the sense in which a recognition of the validity of the pure categories requires the adoption of a 'more than human' perspective. At CPR, B 144-5, Kant sums up the first part of the B-Deduction by referring to a "unity...which the category (according to par. 20) prescribes to the manifold of

to laws of their own. But appearances are only representations of things which are unknown as regards what they may be in themselves. As mere representations, they are subject to no law of connection save that which the connecting faculty prescribes."

⁵⁴⁵ My emphasis in all passages: "Nature, considered merely as nature **in general**, is dependent upon these categories as the original ground of its necessary conformity to law (*natura formaliter spectata*). Pure understanding is not, however, in a position, through mere categories, to prescribe to appearances any a priori laws other than those which are involved in a nature as such, that is, in the conformity to law of all appearances in space and time. Special laws, as concerning those appearances which are empirically determined, cannot in their specific character be derived from the categories, although they are one and all subject to them. To obtain any knowledge whatsoever of these special laws, we must resort to experience; but it is the a priori laws that alone can instruct us in regard to experience **in general**, and as to what it is that can be known as an object of experience." (CPR, B 164-5) This is why the categories are to be regarded, only but crucially, as "fundamental concepts by which we think objects **in general** for appearances" (CPR, A 111).

a given intuition in general." This means that the categories impose *conditions* of unity on the manifold of given intuitions and on the subject of synthesis that is affected with this manifold. The qualification 'in general' is, once more, very significant here. It suggests that the content of the category, and hence the condition it 'prescribes', contains no reference whatsoever to specifically human forms of sensibility. This point is, I think, essential to understanding Kant's notion of 'intellectual synthesis' (Verstandesverbindung). When Kant uses this term, he does not refer to a synthesis of concepts in judgments⁵⁴⁶, but to the representation of a combination and of a resulting unity of given intuitions of which the category is the condition. This representation refers only to the idea that intuitions are *given* to the understanding prior to its combination; it leaves indeterminate how – i.e., in what form, or 'mode' – the sensible manifold is given:

A manifold, contained in an intuition which I call mine, is represented, by means of the synthesis of the understanding, as belonging to the necessary unity of self-consciousness; and this is effected by means of the category. (...)...[this] represented unity of intuition, by which an object is given... always includes in itself a synthesis of the manifold given for an intuition... (...)...since the categories have their source in the understanding alone, independently of sensibility, I must abstract from the mode in which the manifold for an empirical intuition is given, and must direct attention solely to the unity which, in terms of the category, and by means of the understanding, enters into the intuition. (...) But...there is one feature from which I could not abstract...namely, that the manifold to be intuited must be given prior to the synthesis of understanding, and independently of it. How this takes place, remains here undetermined. (CPR, B 144-5)

Here we need to keep in mind (cf. section 4) that Kant thinks that we are capable of representing a synthesis, and the unity it achieves, as something *to be performed or to be achieved*⁵⁴⁷; and we can form this thought even if we abstract from the forms in which the sensible manifold that is to be synthesized is given. If we are representing a synthesis that brings intuitions *in general* to an objective kind of synthetic unity, we are thereby invoking the condition or 'function' of such unity – namely, the pure unschematized category. In this act of thought we are representing what is contained in the category, namely, its conceptual content. Kant thinks that the pure categories *literally contain* the thought of a synthesis of an intuitive manifold in general⁵⁴⁸ or literally contain a unity of an intuitive manifold in general.⁵⁴⁹ This means (I suggest) that a thought that employs a pure category, independently of any reference to specific forms of sensibility, achieves

⁵⁴⁶ Allison 2004, p. 162, p. 177 suggests that the first part of the B deduction concerns the categories only as discursive rules for judgment, whereas the second part concerns them as rules for the synthesis of sensible intuition. The main problem with this suggestion is that the first part of the B deduction concerns precisely *that* synthetic unity of the manifold of intuition *in general* (CPR, B 145) that becomes the synthetic unity of *spatio-temporal* intuition ('of our senses', B 145) in the second part, and since this is not a discursive unity of concepts in judgments in the second part, it is not a unity of concepts in the first part either. Longuenesse suggests that the first and the second part of the B deduction are concerned with, respectively, the contribution of the categories qua discursive forms (exercised in acts of 'analysis') and the categories qua rules of sensible syntheses (2001, p. 110-1). But I do not think that the emphasis in the first part is any less on the *synthetic* unity of apperception (construed as that which is prior to 'my determinate thought', a notion that surely relates to what Longuenesse would consider as 'analysis') than in the second part.

⁵⁴⁷ As a reminder, see CPR B 135: "I am conscious to myself a priori of a necessary synthesis of representations -- to be entitled the original synthetic unity of apperception -- under which all representations that are given to me must stand, but under which they have also first to be brought by means of a synthesis."

⁵⁴⁸ CPR, B 150-1: "...the synthesis which is thought in the mere category in respect of the manifold of an intuition in general...;" "The synthesis or combination of the manifold in them [i.e., in the concepts]...".

⁵⁴⁹ CPR, A 138/B 177: "The concept of the understanding contains a pure synthetic unity of the manifold as such."

the representation of a synthesis of a given manifold that brings this manifold to synthetic unity, *under the direction of the category* (as the condition of that unity). That this representation abstracts from specific forms of sensibility has various consequences. First, via this representation we can see that for *any* intellect that depends for cognition on given intuitions, *regardless of the forms through which those intuitions are received*, synthesis according to the categories must precede analysis: given intuitions must be related, under the direction of the categories, to one another in one consciousness before they can be analyzed into empirical concepts.⁵⁵⁰ But secondly, as long as we leave out of account the specific character of our human forms of sensibility, the representation enabled by the pure categories leaves indeterminate how, precisely, the category directs the original synthesis, i.e., *what* it 'prescribes' as a condition for relating given intuitions to objects. Correspondingly, *what object* given intuitions can represent as a result of being brought to synthetic unity under the categories also remains indeterminate:

The pure [unschematized] concepts of understanding are free from this limitation [i.e., the limitation of being tied to space and time], and extend to objects of intuition as such, be the intuition like or unlike ours, if only it be sensible.... (CPR, B 148) The pure concepts of understanding relate, through the mere understanding, to objects of intuition as such, whether that intuition be our own or any other, provided only it be sensible. The concepts are, however, for this very reason, mere forms of thought, through which alone no determinate object is known. The synthesis or combination of the manifold in them [i.e., in the concepts] relates only to the unity of apperception, and is thereby the ground of the possibility of a priori knowledge, so far as such knowledge rests on the understanding. This synthesis, therefore, is at once transcendental and also purely intellectual. (CPR, B 150)

The representation of intellectual synthesis and of merely intellectual synthetic unity that forms the content of the pure category represents neither a determinate object nor, therefore, a determinate rule for relating intuitions to objects. It does, however, allow for the recognition of a cognitively important principle (the 'a priori knowledge' Kant mentions in the above passage), namely, the principle of the synthetic unity of apperception expounded at the beginning of the B Deduction (CPR, B 137, my emphasis): "The first pure knowledge of understanding... *which also at the same time is completely independent of all conditions of sensible intuition*, is the principle of the original synthetic unity of apperception." This principle states that all given intuitions that can be united in one self-consciousness and that can (thus) be related to objects are subject to the conditions of the synthetic unity of apperception, and hence (cf. par. 19-21) to the pure concepts of the understanding. As Kant indicates, this principle applies to any discursive understanding in complete independence of the conditions of sensible intuitions to which it is subject. This is the sense in which a recognition of the validity of the categories, qua sources of rules for synthesis, requires that one adopts a 'more than human standpoint'.

Of course, we must bring these abstract insights and mere forms of thoughts to bear on our specific forms of sensibility in order to progress from the representation of *objects of intuition in general* to the representation of *objects of human spatio-temporal intuition in general*, and in order to progress from the representation of laws for combining any given manifold to the representation of laws for combining a given spatio-temporal manifold, i.e., for representing a 'unity of the synthesis of the manifold' in space and in time. This latter unity is, Kant tells us (CPR, B 161), "no other than the unity of the combination of the manifold of a

⁵⁵⁰ Every discursive intellect comes equipped with a set of logical forms of judgment and, given the argument of § 19-21 of the B Deduction, from those logical forms a set of categories and a set of rules for the synthesis of intuitions can be derived.

given intuition in general in an original consciousness, in accordance with the categories, in so far as the combination is applied to our sensible intuition". The representation of this unity of space and time puts a priori constraints on everything that is to be determinately represented as a spatio-temporal object, namely, on (CPR, B 159) "whatever objects may present themselves to our senses...in respect of the laws of their combination" – i.e., on empirical objects, or appearances, *in general*, regardless of their specific empirical constitution.⁵⁵¹ Once the merely human forms of intuition are brought into play, and the spatiotemporal schemata are employed to restrict (and objectify) the pure categories, one has clearly left behind a perspective that is available to all discursively thinking beings. Nonetheless, it is a deep fact that the principle, 'for every succession in time, I must posit a causal condition from which it follows necessarily and invariably' (alternatively, 'every alteration in time has a cause') owes its validity to a form of thought that is 'more than human': it is based on the general notion that discursively thinking beings must combine given intuitions according to the concept of causality (in order to have objective thought), and hence it draws on a general form of thought (the pure category of causality) that is shared among all discursively thinking beings. Thinking beings with forms of sensibility that differ from ours would be bound by an analogous principle, involving whatever schema they would use for restricting the pure category of causality to appearances.⁵⁵²

7: The spontaneity of the understanding and the doctrine of noumenal ignorance

I want to conclude by noting the consequences that my line of interpretation has for the self-consciousness and self-knowledge that we exhibit as subjects of spontaneous thought.

Here I want to begin by drawing attention to the fact, which will turn out to be very important for my final chapters, that according to Kant we exhibit the same generic kind of self-consciousness as practical and as epistemic agents: an awareness of ourselves as 'intelligences'

⁵⁵¹ The progression from a less to a more determinate representation of objects of intuition *in general* comes out nicely in this passage (my emphasis): "When...I perceive the freezing of water, I apprehend two states, fluidity and solidity, and these as standing to one another in a relation of time. But in time, which I place at the basis of the appearance [in so far] as [it is] inner intuition, I necessarily represent to myself synthetic unity of the manifold, without which that relation of time could not be given in an intuition as being determined in respect of time-sequence. Now this synthetic unity, as a condition a priori under which I combine the **manifold of an intuition in general**, is -- **if I abstract from the constant form of my inner intuition, namely, time** -- the category of cause, by means of which, when I apply it to my sensibility, I determine everything that happens in accordance with the relation which it prescribes, and I do so **in time in general**. Thus my apprehension of such an event, and therefore the event itself, considered as a possible perception, is subject to the concept of the relation of effects and causes, and so in all other cases" (CPR, B 163-4).

⁵⁵² I want to relate this point to my discussion of the more than human validity of practical laws (chapter VIII, section 5). There, I argued that the principle, 'One ought to donate to Oxfam rather than buy a new Ipod' has a more than human validity despite the fact that it represents the concrete duty through empirical (and hence merely human) concepts: this is because the principle's moral point rests on a practical idea that is binding for all rational beings (roughly, the idea that rational beings have a kind of worth that is incomparably superior to the market price of Iphones). Rational beings with forms of sensibility that differ from ours would represent *the exact same duty* in terms of empirical concepts informed by their forms of sensibility; in this sense, sensibility and the empirical are inessential to the moral principle. I do not think that we can say the same about the epistemic principles that employ the schematized categories. If we focus on the subject-directed aspect of these principles, it seems that the aim to represent a unified *spatiotemporal* order is essential to their epistemic 'point'. Moreover, unlike practical laws, the principles of the understanding also have an object-directed aspect: they yield of cognitions of empirical objects in general. This aspect strictly requires that the pure categories (which are mere forms of thought) are supplemented with the schemata and hence imbued with spatiotemporal content.

belonging to an 'intelligible' world rather than as phenomenal objects belonging to the empirical world disclosed by our senses. Compare the following passages (my emphases):

Even as to himself, a man cannot pretend to know what he is in himself from the knowledge he has by internal sensation. (...) Thus in respect to mere perception and receptivity of sensations he must reckon himself as belonging to the world of sense; but in respect of whatever there may be of pure activity in him (that which reaches consciousness immediately and not through affecting the senses), he must reckon himself as belonging to the intellectual world, of which, however, he has no further knowledge. (...)...a rational being must regard himself qua intelligence (not from the side of his lower faculties) as belonging not to the world of sense, but to that of understanding.... (...) **Every rational being reckons himself qua intelligence as belonging to the world of understanding, and it is simply as an efficient cause belonging to that world that he calls his causality a will.** (GMS, 4: 451-2)

...so far as inner intuition is concerned, we know our own subject only as appearance, not as it is in itself. On the other hand, in the transcendental synthesis of the manifold of representations as such, and therefore in the synthetic original unity of apperception, I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am. This representation is a thought, not an intuition. (...) **I exist as an intelligence which is conscious solely of its power of combination.** (...) **Such an intelligence**, therefore, can know itself only as it appears to itself in respect of an intuition which is not intellectual and cannot be given by the understanding itself, not as it would know itself if its intuition were intellectual. (CPR, B 157-9)

Kant holds that we are aware of ourselves as 'an intelligence' and as a member of a non-sensible world when we exercise our 'higher' practical *and* theoretical faculties, that is, our volitional and our epistemic powers (respectively, of choice or 'of combination').⁵⁵³ Now, it must be admitted that Kant is not perfectly consistent about the *epistemological status* of this self-consciousness. His official line of thought, which is prevalent in both the above passages, is that we *know* ourselves only as members of the sensible world. This is because one can know anything as it is only through intuition, and our only intuitive resources are sensible; in order to know ourselves as we are in ourselves, as members of an intelligible world, we would have to know ourselves through an intellectual intuition, which is impossible for beings like us. As Kant puts it:

Such an intelligence, therefore, can know itself only as it appears to itself in respect of an intuition which is not intellectual and cannot be given by the understanding itself, not as it would know itself if its intuition were intellectual. (CPR, B 159)

However, in an important passage from the Dialectic (CPR, A 546/B 574), Kant states that "man...knows himself also through pure apperception", rather than "solely through the senses", and "is thus to himself ...a purely intelligible object" in respect of "the action of... faculties", such as the understanding, whose exercise is empirically unconditioned (spontaneous).

So, as interpreters, we are faced with the problem that Kant makes apparently conflicting claim about the epistemological status of our self-awareness as spontaneous beings. If we press the issue further, it seems that we are faced here with a dilemmatic choice. On the one hand, it seems (i) that Kant is clearly overstepping the bounds of his critical epistemology (delineated by his thesis of noumenal ignorance) when he declares (in the Dialectic) that we can know our intelligible character. But on the other hand, it seems (ii) that simply denying that we

⁵⁵³ See chapter VIII, section 4 for my suggestion that we should further differentiate this self-awareness into (a) a consciousness of ourselves as purely spontaneous, i.e., as capable of producing representations in complete independence from the deliverances of our receptive faculties, and (b) a consciousness of ourselves as both spontaneous and receptive, as members of both the sensible and the intelligible world.

can know our intelligible character not only forces us to dismiss the above Dialectic passage: it also leaves us wondering what Kant could mean in the many passages where he stresses that we are conscious of ourselves as an 'intelligence' that belongs to the non-sensible world.

Before trying to offer a response to this dilemma, I want to clarify what is at stake in (i). Notice that the claim *that we have* an intelligible character that differs from our sensible, phenomenal or empirical character is completely unproblematic within the confines of Kant's critical idealism. This claim follows from the claim – which Kant thinks he can establish conclusively – that the spatiotemporal features that make up our sensible character do not exist apart from our minds (cf. my chapter I). Similarly, the claim that things in themselves are not in space in time does not violate the thesis of noumenal ignorance because it is a merely negative statement that does not amount to 'proper knowledge' (CPR, B 149; see my chapter I, section 1). The problem to which (i) calls attention is, rather, that in the Dialectic passage Kant is saying that we know that our intelligible character comprises the capacity for spontaneous, sensibly or empirically unconditioned activity. This claim is problematic because it clearly goes beyond the thought that we have some indeterminate kind of noumenal character that differs from the character we have qua appearances.⁵⁵⁴ Here it is important to notice that in the Dialectic passage Kant clearly suggests that we know *more* about the noumenal character of human agents than about that of other natural objects, such as animals or inanimate matter:

In lifeless, or merely animal, nature we find no ground for thinking that any faculty is conditioned otherwise than in a merely sensible manner. Man, however, who knows all the rest of nature solely through the senses, knows himself also through pure apperception; and this, indeed, in acts and inner determinations which he cannot regard as impressions of the senses.

This cannot mean that animals and inanimate objects of nature do not have an ontological character that differs from their phenomenal nature. What Kant must mean here is that it is only with regards to human beings that we have grounds for asserting what their noumenal or intelligible character must be like, namely, that it is to be characterized in terms of a capacity for spontaneous activity that is not determined or conditioned by sensible causes. And this does seem hard to square with the doctrine of noumenal ignorance.

I do not pretend to have a fully satisfactory solution to this problem, but I think that there are a few non-arbitrary distinctions that we can avail ourselves of to minimize its force. One such distinction can be gathered from an important footnote in the Deduction:

Now since I do not have another self-intuition which gives the determining in me (I am conscious only of the spontaneity of it) prior to the act of determination, as time does in the case of the determinable, I cannot determine my existence as that of a self-active being; all that I can do is to represent to myself the

⁵⁵⁴ This is a sympathetic way of understanding Walker's worry (1978, pp. 131-5). Walker himself frames the worry in terms of the question, "But how can I presuppose it [the noumenal self] without knowing something about it, namely that it exists?" (p. 132). As I said, it seems unproblematic for Kant to claim *that*, for any mind, the mind exists as a noumenal mind (or self). What Walker goes on to say suggests that it is really the appeal to the self-active, spontaneous nature of our noumenal self that seems troubling. Against Ameriks, I do not think that this worry can be defused by appeal to the idea that Kant's Idealism requires the notion that "there is some absolute reality responsible for synthesis" (2000, pp. 72-3). Walker's worry, as I understand it, pertains to the idea that we know what the subject responsible for such synthesis *is like*. When Ameriks claims (ibid., p. 287) that "we have no ground for asserting that there is an 'agent' here", he is ignoring Kant's repeated insistence that we are aware of ourselves as *self-active* beings, in respect of faculties whose actions cannot be conceived as determined by empirical or sensible causes (see, again, the above passage from the Dialectic).

spontaneity of my thought, that is, of the determination; and my existence is still only determinable sensibly, that is, as the existence of an appearance. But it is owing to this spontaneity that I entitle myself an intelligence. (CPR, B 157-8)

The distinction I have in mind is between (a) the idea that 'we can recognize ourselves as having a certain property' and (b) the idea that 'we are capable of determining ourselves as having that property'. The relevant property is that of spontaneity, i.e., it concerns a mode of existence that Kant often (as in the above passage) refers to as 'self-activity'. This is precisely the property that the problematic Dialectic passage is concerned with: the capacity for activity that is not empirically or sensibly conditioned. I suggest that Kant's claim that we know ourselves as spontaneous noumena or intelligible beings (in the Dialectic) and Kant's claim that we are conscious of ourselves as intelligences (in the Deduction and the Groundwork) respects the commitment to noumenal ignorance at least insofar as it denies (b). While we know (a) *that we have* the non-empirical/non-sensible capacity for spontaneous activity, we cannot determine our existence as a self-active being: that is, we have no theoretical insight into how this capacity works.⁵⁵⁵ Our conception of ourselves as being capable of spontaneous activity is theoretically fruitless in that we cannot observe, explain, understand or predict the workings of this capacity. (This fits with how Kant puts things in the Groundwork: we must 'reckon ourselves' as belonging to the intelligible world by virtue of our capacity for spontaneous activity, but we have "no further knowledge" of ourselves qua non-sensible object.) The spontaneous activity of the thinking self cannot become an object of theoretical knowledge: it cannot be determinately understood just like (say) the empirical self's combination of representations according to the law of reproductive association can be understood, or (even more so) like the movement of bodies according to the law of gravity can be understood. These considerations are, in my view, of a piece with Kant's insistence that "the subject of the categories cannot by thinking the categories acquire a concept of itself as an object of the categories" (CPR, B 422).

Moreover, we can give further substance to the denial of (b). A crucial point of Kant's insistence that we cannot 'determine' our existence as a self-active being is that an appeal to our consciousness as a self-active being cannot serve as a ground of insight into the metaphysical constitution that we possess as a self-active (spontaneous) being. Our consciousness as a spontaneous intelligence does not, by itself, support *further* inferences to the simplicity, the immortality, the personal identity over time, or even the immateriality of the noumenal self.⁵⁵⁶ All

⁵⁵⁵ This distinction is, in my view, at least part of what Kant has in mind when he makes the notoriously obscure remark that in transcendental synthesis "I am conscious of myself, [1] not as I appear to myself, nor [2] as I am in myself, but [3] only that I am" (square brackets my addition). [1] is unproblematic; it is clearly intended to rule out the idea that our awareness of ourselves as subjects of synthesis pertains to our phenomenal, empirical mode of being (i.e., to the way we are qua appearances). But what does Kant intend with the distinction between [2] and [3]? Part of the point must be that qua subject of synthesis I am conscious that I am a thing in itself (since, via [1], I am not conscious of myself as an appearance) without being able to determine my existence as a thing in itself. But clearly this is not sufficient: Kant is committed to saying that qua subject of synthesis I grasp not only that I am a thing in itself but also that I am a spontaneous agent of thought, an intelligence. Still, with the distinction I suggested, there is a sense in which I am not conscious of myself *as* I am in myself, since I cannot possibly answer any questions as to *how* my capacity for spontaneous thought operates.

⁵⁵⁶ Kant argues for the (theoretical) unknowability of the noumenal self in all these senses in his discussion of the Paralogisms of Pure Reason. The claim that we do not know whether the noumenal self is *immaterial* has a special status: Kant must admit that we know that the noumenal self is non-material in the sense that it cannot be characterized in terms of spatial predicates (because space is only a form of intuition and the constitution of

these crucial tenets of the rationalist doctrine of the soul are theoretically unknowable in Kant's critical philosophy, and this central feature of his critical epistemology survives the concession that we do know *that* the noumenal self is capable of a purely spontaneous activity.

One might insist that these considerations cannot alleviate the worry that Kant is overstepping the bounds of his critical epistemology when he claims that we know, on the basis of our consciousness of ourselves in acts of synthesis, that we have a non-sensible capacity for spontaneous (self-)activity. In response to this worry, I want to introduce a further distinction between two sources of self-knowledge. First, one might know that one has certain properties on the basis of intuition, observation, or introspection. Kant is committed to saying that on this basis we can only know ourselves as phenomena; if he did claim that we can consider ourselves as purely spontaneous noumena on the basis of some kind of intuition or introspection, his position would collapse into dogmatic rationalism. But there is a second potential source of self-knowledge: our grounds for ascribing a certain capacity to ourselves might be that it is only in virtue of possessing this capacity that we can have knowledge of objects. This, I suggest, is Kant's 'critical', non-dogmatic ground for claiming that we know that we are spontaneous intelligences: while our capacity for spontaneous activity (e.g., thought) cannot itself become an object of theorizing, we must presuppose our possession of this capacity in all our theorizing about objects. Hence, we can infer to our spontaneity on the basis of the idea that such spontaneity is a precondition for all our theoretical, empirical knowledge.

This line of thought crucially depends on the idea that we cannot consider our capacity for synthesis of intuitions as a sensible capacity that pertains to our phenomenal self and whose exercise is governed by empirical laws. It should be clear (pace Kitcher) that Kant affirms this idea: this follows from his claims that as subjects of synthesis we are not conscious of ourselves as we appear (CPR, B 157) and that "through pure apperception" we know that some of our actions are the result of empirically unconditioned capacities (CPR, A 546/B 574). But one might wonder why he affirms this idea. Why does he hold that in our empirical knowledge claims we must presuppose an empirically unconditioned capacity for spontaneous thought?

I want to explain Kant's point in three steps. First, he holds that any legitimate empirical knowledge claim presupposes the validity of the categories. Second, he holds that we can only affirm the validity of the categories if we can show that they apply necessarily to objects of experience, that is, if certain principles about empirical objects are necessarily true (e.g., 'every natural alteration as such has a cause'). Third, he holds that we can only affirm the necessity of the principles that employ the categories, or the necessary applicability of the categories, if we can affirm that the categories are produced spontaneously by the empirically unconditioned understanding.

The first step rests on the following idea: the validity of the categories must be presupposed in empirical knowledge claims because the categories are conditions for the possibility of representing public, mind-independent objects of experience, and this is a basic precondition for any less basic, more sophisticated claim about the course of nature.

The second step is to be defended as follows; I shall focus here on the concept of causality. This concept expresses that an event follows necessarily from a preceding event. That

noumena is independent of such forms). There is, however, a remaining sense in which the non-materiality of the noumenal self is unknowable: we cannot know whether the noumenal reality that underlies appearances of mere inner sense (the mental) is identical to or different from the noumenal reality that underlies appearances of outer sense. See CPR, B 427-8; Ameriks 2000, p. 34 calls this the unknowability of 'Transcendental Immaterialism'.

two events are connected in terms of empirical necessity (cf. CPR, A 228/B 280-1) requires empirical observation, but it cannot *merely* be established on this inductive basis: a necessary connection among events is not a possible object of intuition.⁵⁵⁷ Hence, establishing that the concept of causality applies to experience at all requires that one can show that it applies to objects of experience necessarily: it is only if one can show that thinking in terms of this concept is a necessary condition of empirical thought that one can demonstrate its validity.⁵⁵⁸

The third step involves an inference from the necessity of certain principles to what Kant sometimes calls the 'purity' of the faculty that is responsible for these principles. Kant explicitly draws this inference at KpV, 5: 30 (a passage that will play a central role in my argument in chapter XIII). To say that a faculty is 'pure' is to say that it contains pure, a priori representations or principles which do not have an empirical origin. We already saw (cf. chapter IX, section 6) that Kant holds that a principle that expresses a strict necessity would lose its claim to validity if it could be shown that the principle has its origin in our contingent empirical, psychological constitution (cf. EEUK, 20: 238). His point here, I suggest, is that if we had to concede that our thinking in terms of a certain concept *x*, whose content is not displayed in sense perception, is determined by our contingent psychological (or, as we might say, neurological) constitution, then we could not take ourselves to be rationally entitled to thinking in terms of *x*. We would have to agree that our thinking in terms of *x* is the mere upshot of our empirical minds or brains being wired in an arbitrary manner.⁵⁵⁹

I will scrutinize and, to some extent, motivate the ideas inherent in this third step in chapter XII. For present purposes, let us suppose that the line of thought displayed in these three steps is cogent. And now suppose someone – a committed naturalist – were to respond that *if* this line of thought is indeed compelling, we should jeopardize the validity of the categories rather than accept that the thinking mind that produces the categories is not determined by empirical causes. This naturalistic response claims that our entitlement to the notion that all our activity is governed by the causality of nature is more securely grounded than our entitlement to the notion that some of our activity must be spontaneous in the sense of being independent from natural causality. But notice now that from Kant's point of view, this response is wholly unsatisfying: in

⁵⁵⁷ "...no one will say that a category, such as that of causality, can be intuited through sense" (CPR, A 137-8/B 176-7); no *singular* empirical (or, indeed, pure; see CPR, A 732-3/B 760-1) intuition can exhibit the concept of empirical necessity (see also Guyer 1987, pp. 164-6). A *repeated* observation of event patterns will by itself only establish what Kant calls the 'comparative universality' of the pattern (CPR, B 3-4). This does not mean that a repeated observation of an event pattern cannot provide us with empirical evidence for there being a genuinely necessary connection, but this *presupposes* that the concept of a necessary connection between events has already been legitimized; this legitimization is impossible on the basis of empirical observation.

⁵⁵⁸ Notice that there are two distinct modalities here: the notion that natural events are connected through empirical, nomological necessity (n1) is grounded in the entirely different modality that one invokes when one claims that we must necessarily (n2) think in terms of the concept of causality. Notice, further, that the necessity at issue in (n2) is not analytic necessity: there is no contradiction in the conception of events following each other randomly. The concept of an alteration or event implies that whenever something happens in time, something else precedes it, but it does not imply that two events in time are necessarily connected as cause and effect (CPR, A 9/B 13). Likewise, the concept of cause itself implies that whenever the cause occurs, the effect must also occur, but again this cannot justify the claim that two events *are* actually connected as cause and as effect.

⁵⁵⁹ Hence Kant's rejection of the idea that our use of the concept of cause rests on an "arbitrary subjective necessity" (CPR, B 167-8); or his corresponding claim that in making causal claims about nature one must presuppose "that the understanding has the capacity to determine its judgment according to objective reasons that are valid at any time, and does not stand under the mechanism of merely subjectively determining causes" (RezSchulz, 8: 14).

supposing that all our activity is governed by natural causes, the committed naturalist simply takes for granted the validity of the concept of causality, which is one of the categories whose validity the naturalistic responses purports to dismiss. That is, the naturalist is supposing that the concept of causality has validity even though, on a naturalistic worldview, our thinking in terms of this concept is (since the content of this concept is not displayed in sense perception) the mere upshot of our arbitrary psychological (or neurological) constitution. For Kant, the legitimacy of the appeal to a (deterministic) causality of nature that reigns universally within the phenomenal world presupposes the appeal to a kind of (non-phenomenal) activity that is not itself governed by such causality. The naturalist who claims that all causality is nature, and (hence) that all thinking activity stands under natural causes, is pulling the rug out from under her own feet.

This is how I envisage Kant's defense of the claim, 'We know that our non-phenomenal mind (or understanding) comprises the capacity for spontaneous epistemic activity'. His defense of this claim is not based on an appeal to some kind of intuitive insight, but on the reflection on the conditions of possible experience, i.e., the legitimacy and strict necessity of those principles and concepts in terms of which we must think in order to have experience of the empirical world. Hence, Kant is not committed – not even by his strong remarks in the *Dialectic* – to the notion that the subject of thought has immediate consciousness of herself as a spontaneous, noumenal being or object. What we are *immediately* aware of as subjects of synthesis is the determination of our thought by pure, a priori rules or concepts. It is on the basis of this self-consciousness that we can infer to our noumenal capacities as the seat of these rules (albeit, to repeat, this does not give us any further metaphysical insight into how we are constituted as 'intelligences' or as 'intelligible objects'). Hence, this defense of the epistemic spontaneity of our noumenal self has a unique basis (i.e., in rests on an appeal to what we must suppose about our noumenal self in order to claim to know anything at all) that cannot be extended to other putative properties of our noumenal self (such as immortality, etc.): it does not, pace Ralph Walker, 'open the flood-gates' for rampant metaphysical theorizing about how we are in ourselves.

In particular, it should be noted that the exact same line of thought *cannot* be invoked to defend the claim, 'We know that our non-phenomenal will comprises the capacity for spontaneous practical activity'. This is because the concepts that constitute the conditions of objective experience of the empirical world (and whose legitimacy must hence be presupposed in claims about the causal conditions that govern natural events) do *not* include the concepts that are employed in the principles established by *practical reason*. The categorical imperative is, in contrast to the causal principle, completely irrelevant for experience of natural objects from the empirical standpoint. Hence, while Kant thinks that one *can* argue from our experience and knowledge of empirical objects to the purity/spontaneity of the understanding qua source of the causal principle (a principle which is a condition for the possibility of all empirical knowledge), he does *not* think that one can argue from the experience and knowledge of empirical objects to the purity/spontaneity of practical reason qua source of the categorical imperative (a principle which cannot be considered a condition for the possibility of empirical knowledge). This suggests another sense in which Kant's appeal to the epistemic spontaneity of our noumenal self is modest: it has no direct implications for the spontaneity of our *will* or *practical reason*.

I will address the question of whether the appeal to epistemic spontaneity can provide a more indirect defense of our idea of practical spontaneity in chapter XIII.

Chapter XI: Two Types of Epistemic Norms

In the preceding chapters, I argued that it is a significant aspect of Kant's epistemology that our thinking about nature has a reflective, norm-governed dimension. In chapter IX, I examined his conception of the normative dimension of conceptually determinate empirical judgments. In chapter X, I examined his conception of the normative dimension of the cognitively more basic synthesis of intuitions that epistemically precedes empirical judgement – and concept-formation. In this chapter, I examine the types of epistemic norms that govern our epistemic agency at each of these stages. This requires a discussion of Kant's famous contrast between constitutive and regulative epistemic principles. I shall argue that construing these principles as two distinct types of epistemic norms provides a key toward understanding Kant's contrast and the significance that this contrast has within his critical philosophy.

1: Constitutive and regulative principles: a first approximation

I want to begin my discussion by noting that Kant draws the contrast between constitutive and regulative epistemic principles *twice over*. He asserts, with regards to the laws of the understanding, that the mathematical principles (that employ, chiefly, the categories of quantity and quality) are constitutive whereas the dynamical principles (that employ, among others, the categories of substance and causality) are merely regulative (CPR, A 179-80/B 222-3); but he also asserts that the dynamical principles are constitutive whereas the principles of theoretical reason are merely regulative (CPR, A 664/B 692).

Part of the solution to this puzzle is that the mathematical principles are constitutive of perception⁵⁶⁰, whereas the dynamical principles are constitutive *of experience* (qua necessary combination of perceptions in one consciousness) only; the principles of theoretical reason are not constitutive of anything. But this leaves us with the interpretive challenge to specify a sense in which the dynamical principles, which are *both* constitutive and regulative, are like and unlike the (merely constitutive) mathematical principles and the (merely regulative) principles of theoretical reason. I want to begin by specifying in a programmatic way the general point of Kant's constitutive-regulative taxonomy; this will be made more precise in following sections.

The mathematical principles state that all appearances have an extensive and an intensive magnitude. When we are affected by objects, we receive an intuition that, if it reaches consciousness, becomes a perceptual representation; and we know of the object of this representation that it has an extensive and an intensive magnitude (its intensive magnitude being the real quality in the object that corresponds to our sensation of it). Moreover, with respect to any given object we can show, with mathematical certainty, what its precise extensive and intensive magnitudes is. This latter point is decisive for our purposes: Kant calls the mathematical principles constitutive *because* they are susceptible to the method of mathematical construction, which puts us in a position to determine in an a priori manner, *without appeal to any further experience*, the concrete way in which a given perception instantiates the mathematical categories (i.e., the precise extensive and intensive magnitudes of appearances).⁵⁶¹

⁵⁶⁰ Perception is a representation that is conscious but does not stand under the objective unity of consciousness. See CPR, B 207; see also footnotes 514 and 515 above (in chapter X).

⁵⁶¹ CPR, A 178-9/B 221: "The two previous principles, which, as justifying the application of mathematics to appearances, I entitled the mathematical.... (...) Both principles justify us in employing numerical magnitudes, and so enable us to determine appearance as magnitude. For instance, I can determine a priori, that is, can

The Second Analogy (the paradigm case of a dynamical principle) says that "all alterations take place in conformity with the law of the connection of cause and effect" (CPR, B 233). For any change, we know that it is necessitated by some cause whose causality is governed by an empirical law. We know with a priori certainty that there must be *some such* necessitating empirical ground, but we cannot indicate with a priori certainty *what* this ground is, or what specific empirical law governs the causality of the cause. So the Second Analogy is *like* the mathematical principles in that it tells us that things in nature must be a certain way; it is *unlike* the mathematical principles in that finding out how the category of causality is instantiated requires a posteriori methods and standards, and hence a recourse to experience.⁵⁶²

The "regulative law of systematic unity" (the paradigm case of a principle of theoretical reason) "prescribes that we ought to study nature as if systematic and purposive unity, combined with the greatest possible manifoldness, were everywhere to be met with, in infinitum" (CPR, A 701/B 729). This law is *like* the Second Analogy in that it is rule for positing something within the course of nature, but there is a decisive difference: the Second Analogy entitles us to *knowledge* about nature, that is, it provides us with firm and unassailable objective grounds for assenting to the judgment that for any given alteration there must be a causal ground that necessitates the occurrence of that alteration according to a universal rule. By contrast, the law of systematic unity does not entitle us to know that the empirical laws and concepts that we discover will systematically and purposively combine with other empirical laws and concepts.

1: The Analogies as constitutive epistemic norms

I now want to explain in greater detail the sense in which the dynamical principles, and in particular the Analogies, are constitutive. The constitutive character of these principles has two interconnected sides, which we may refer to as a *subject-directed* and an *object-directed* aspect. The subject-directed aspect of constitutive epistemic principles captures their distinctively normative dimension, i.e., it signals that those principles normatively prescribe an act of thought that is necessary for the subject of thought to get into the epistemically significant state of *experience* of the empirical world. A violation of such a norm signals that the subject is caught up in the wholly subjective play ('rhapsody', 'dream') of her representations (see CPR, A 156/B 195; and A 194/B 239), which fail to relate to mind-independent objects or to an intersubjectively accessible, public world. In this sense, compliance with the dynamical principles is constitutive of experience. The object-directed aspect of these principles consists in the fact that they give us (synthetic a priori) knowledge of the world of appearances (i.e., of nature). This aspect comes to light in Kant's tendency to formulate these norms in a way that leaves out any reference to the

construct, the degree of sensations of sunlight by combining some 20,000 illuminations of the moon. These first principles may therefore be called constitutive."

⁵⁶² Guyer 1987, pp. 187-8, pp. 201-2 argues that there is no relevant difference between the mathematical and the dynamical principles: the mathematical principles – like the dynamical ones – entitle us only to some indeterminate conclusions, namely, to the conclusion that appearances instantiate *some* degree of intensive or extensive magnitude whose precise number must be determined empirically (just like the precise instantiation of the dynamical categories must be determined empirically); hence they, too, must be considered regulative. But even if Guyer is right that the determination of the relevant magnitudes requires "methods of laboratory experiments", Kant can still insist that the experimental methods for determining the empirical instantiation of the mathematical categories have an a priori underpinning (construction in pure intuition; see the preceding footnote) and, very importantly, that (in part because of this a priori foundation) our methods are guaranteed cognitive success. None of these points apply in the case of the dynamical categories.

subject as cognition, as claims about the world: "All alterations take place in conformity with the law of the connection of cause and effect". I have explained this crucial part of Kant's position in some detail in chapter X (see sections 4 and 7): for Kant, the concepts employed in the principles of the understanding are constitutive not just of a thinker's experience of objects but also *of the objects* thus experienced: "the conditions of the possibility of experience in general are at the same time conditions of the possibility of the objects of experience" (CPR, A 158/B 197). The objects of knowledge depend, for their character, on our faculties of knowledge. The normative principles that elucidate necessary conditions for the experience of objects are (one might say) 'ontologically transparent' to the way objects (as appearances) necessarily must be.

In order to properly understand these ideas, we must emphasize the *indeterminacy* of the Analogies, which has implications both for their subject-directed and for their object-related aspect. Here it will be illuminating to contrast my interpretation of the constitutive character of the dynamical principles with that of Michael Friedman. Friedman's account involves two claims: first (I), the constitutive character of the dynamical principles derives from the fact that they make possible (in an a priori fashion) certain empirical concepts, such as alteration, without which no experience takes place.⁵⁶³ Second, (II), Kant engages in the project of showing how the dynamical principles make empirical concepts possible in the *Metaphysical Foundations of Natural Science*, where (for instance) the Second Analogy is determined into the principle that every alteration *of state of motion* is caused by a moving force.⁵⁶⁴

I do not think that there is much of a textual basis for (I).⁵⁶⁵ While Kant indeed says that the concept of alteration that figures in the Second Analogy is drawn from experience⁵⁶⁶, he never quite says that *this* concept is one without which no experience takes place: the concept which is a condition of the possibility of experience is the a priori concept of causality. This concept is made possible, or validated, by the principle, 'every alteration has a cause'.⁵⁶⁷ Now Friedman might concede this and say that all that matters to his interpretation is that the concept of alteration plays an indispensable role, via its contribution to the content of the Second Analogy, in making experience possible. This, I think, is a correct and important point. But it does not establish (II); and seeing why it does not establish (II) is of crucial significance for understanding the exact sense in which the Analogies are constitutive both of experience and of

⁵⁶³ Friedman 1992c, pp. 79-80.

⁵⁶⁴ Friedman 1992c, p. 81.

⁵⁶⁵ Friedman's defense of (I) draws on a passage at CPR, A 664/B 692: "(...) the dynamical laws in question are... constitutive with respect to experience, in that they make the concepts without which no experience takes place possible a priori. The principles of pure reason, by contrast, cannot even be constitutive with respect to empirical concepts, because no schema of sensibility can be given corresponding to them, and they therefore have no object in concreto." Friedman suggests (1992c, p. 79) that Kant's appeal to empirical concepts in the second sentence obliges us to think that the concepts mentioned in the first sentence must also be empirical. But this does not accord with Kant's rhetoric: after noting that the dynamical laws owe their constitutive character to the fact that they make possible concepts that are necessary for experience, he says that the principles of pure reason by contrast cannot *even* make empirical concepts possible. And this strongly implies that the concepts he mentions in the first sentence must be a priori concepts: there would be no explanation for the 'even' if he was referring to empirical concepts already in the first sentence.

⁵⁶⁶ See CPR, B 3-5. The fact that the Analogies contain an empirical concept renders them *impure* synthetic a priori principles; the contrast is, again, with the mathematical principles, which are pure synthetic a priori principles because they do not contain empirical concepts.

⁵⁶⁷ That is: the (proof of the) principle that we could not objectively represent an alteration without thinking of it as falling under the concept of causality validates the concept of causality.

objects. The concept of alteration that figures in the Analogies is more general than the concept of alteration that figures in the argument(s) of the Metaphysical Foundations: the latter concerns only a specific kind of alteration, namely, alteration of *state of motion*; and the constitutive character of the Analogies cannot be restricted in this manner. Here Friedman's interpretation does not accord with one of Kant's main ambitions in the First Critique, namely, the ambition to secure that we know that *every* natural alteration as such has a necessitating causal ground. If the constitutive character of the Analogies did depend on the role that they play within the argument of the Metaphysical Foundations, this ambition would have to be given up, because this argument concerns only *part* of nature: namely, objects of outer sense, and indeed only objects of outer sense as they figure in fundamental physics. But Kant holds that the Second Analogy extends to the 'inner' experience of 'the soul'.⁵⁶⁸ Hence the scope of the Analogies, as principles that are constitutive of objects and experience of objects, must extend beyond the focus on fundamental physics that is characteristic of the argument of the Metaphysical Foundations.

What this shows is that the Analogies are 'metaphysically indeterminate', in the sense that they imply nothing about the concrete manner in which the a priori concepts that they employ are instantiated. This indeterminacy is a direct function of Kant's ambition to defend the Analogies as laws of nature *in general, überhaupt*. They provide us with abstract information about what *every* empirical object, or appearance, must be like without supporting, prior to experience, any specific determination of this general empirical character (which requires empirical concepts):

Categories are concepts which prescribe laws a priori to appearances, and therefore to nature, the sum of all appearances. (...) **Nature, considered merely as nature in general**, is dependent upon these categories as the original ground of its necessary conformity to law. Pure understanding is not, however, in a position, through mere categories, to prescribe to appearances any a priori laws other than those which are involved in a **nature in general**, that is, in the conformity to law of all appearances in space and time. Special laws, as concerning those appearances which are empirically determined, cannot in their specific character be derived from the categories, although they are one and all subject to them. (CPR, B 164-5; my emphasis) ...in philosophy the analogy is not the equality of two quantitative but of two qualitative relations; and from three given members we can obtain a priori knowledge only of the relation to a fourth, not of the fourth member itself. The relation yields, however, a rule for seeking the fourth member in experience, and a mark whereby it can be detected. (CPR, A 179-80/B 222)

The final sentence of the second passage quoted here shows that Friedman is correct in saying

⁵⁶⁸ For Kant's commitment to the notion that mental or psychological phenomena stands under the Second Analogy, see (for instance) KpV, 5: 94 ff. I argue for this point in some detail in chapter II, section 2. One might respond, on Friedman's behalf, that Kant thinks that inner experience is dependent on outer experience (CPR, B 276-8). But Kant's 'Refutation of Idealism' shows only that inner experience presupposes a permanent and (given the lack of permanence in mere inner sense) hence an outer reference point for objective time determination. It does not follow that inner experience *is* outer experience, or that alterations of the thinking (empirical) self are governed by the laws that govern alterations of corporeal nature. In fact Kant is quite clear in the Metaphysical Foundations (cf. Man, 4: 467) that the thinking self and body constitute two distinct genera ("Gattungen") of what he calls "nature in the material sense", namely, the epitome of objects of experience. Moreover, he thinks (cf. CPR, 683-4/B 711-2) that empirical laws of corporeal nature are 'of a wholly different kind' than laws that explain mental phenomena (such as the law of reproductive association). Hence, we must expect that the enunciation of principles of corporeal nature leaves indeterminate the principles that govern changes of state of the thinking (empirical) self. Even if we confine ourselves to the domain of corporeal nature, it is clear that the principles of physical science that Kant enunciates in the Metaphysical Foundations capture only part of our 'outer experience' (chemical phenomena are examples of changes of state of body that, as Kant notes, are not obviously governed by fundamental physical laws of repulsion and attraction).

that the Analogies have a role to play in guiding those cognitive pursuits that eventually lead us to the cognition of specific empirical substances, laws, and causal powers. But the crucial point we need to keep in mind is that such determinate cognition of objects is established *on the basis of experience*, whereas the Analogies are principles that make experience possible in the first place. The *only* thing we know about the empirical world by virtue of bringing perceptions under (say) the concept of causality (in accordance with the Second Analogy) is that every successive alteration *as such or in general* is caused by some prior event in accordance with an invariable, necessitating natural law.⁵⁶⁹ This, by itself, affords us with no knowledge whatsoever about what event caused some given alteration, what causal powers are exercised in the production of this alteration, and what causal law governs the alteration.

Since the metaphysical (object-directed) aspect of the Analogies corresponds to their epistemic (subject-directed) aspect, the high degree of indeterminacy in the metaphysical implications that these principles have for natural objects must be mirrored by an equally high degree of indeterminacy in the kind of thinking about natural objects that is governed by these principles qua norms of cognition. Just as the Analogies do not account for the objective reality of any specific empirical concepts or laws (the knowledge that they give us concerns only nature *überhaupt*), so the act of thinking they govern, as norms of synthesis (cf. my chapter IX, section 4), is also prior to the formation or application of specific empirical concepts or laws: it need not (although it may) involve any more conceptual resources than those required for relating given intuitions to empirical objects *in general* and hence for representing an intentional object of thought that is distinct from one's private mental states. The kind of representation thus enabled is called, by Kant, experience, or sometimes also experience *in general* (*überhaupt*). This latter term establishes the congruence between the object-related and the subject-related aspect of the Analogies: just as the synthesis they direct enables (only) experience *in general*, so the implications they have for the way objects are (indeed, must be) concern (only) the character of objects of experience *in general*:

Pure understanding is not, however, in a position, through mere categories, to prescribe to appearances any a priori laws other than those which are involved in a **nature in general**, that is, in the conformity to law of all appearances in space and time. Special laws, as concerning those appearances which are empirically determined, cannot in their specific character be derived from the categories, although they are one and all subject to them. To obtain any knowledge whatsoever of these special laws, we must resort to experience; but it is the a priori laws that alone can instruct us in regard to **experience in general**, and as to what it is that can be known as an object of experience. (CPR, B 165; my emphasis.)

This interpretation can be supported by taking into account how Kant frames the content of the thought that allows us to have objective experience of, for instance, an alteration, i.e., of something that happens (my emphases):

If, then, we experience that something [etwas] happens, we in so doing always presuppose that **something [irgend etwas]** precedes it, on which it follows according to a rule (CPR, A 195/B 240). When, therefore, I perceive that something happens, this representation first of all contains [the consciousness] that there is **something** preceding, because only by reference to what precedes does the appearance acquire its time-relation, namely, that of existing after a preceding time in which it itself was not. But it can acquire this determinate position in this relation of time only in so far as **something** is presupposed in the preceding state upon which it follows invariably, that is, in accordance with a rule.

⁵⁶⁹ This way of putting it presupposes the so-called 'strong reading' of the Second Analogy, which I think (in agreement with Friedman) is the correct one (see chapter I, footnote 79). But nothing hinges on this point here.

(...) The situation, then, is this: there is an order in our representations in which the present, so far as it has come to be, refers us to **some preceding state** as a correlate of the event which is given; and though **this correlate is, indeed, indeterminate**, it none the less stands in a determining relation to the event as its consequence, connecting the event in necessary relation with itself in the time-series. (CPR, 198-9/B 243-4; my emphasis)

It is the experience of objective succession (alteration) in general that is enabled by complying with the Second Analogy, that is, by the mental act of positing ("setzen") an indeterminate *something* from which the alteration follows according to an equally indeterminate necessitating law. The (conceptual, cognitive) *indeterminacy* of what we are positing does not prevent this act of positing from *determining* an object of experience, namely, an objective succession or alteration. The act of positing some causal ground achieves such determination even if we cannot (yet) further determine what this ground is and what law governs its causality.

What this means is that for Kant, *experience*, at a very basic and fundamental level, is not dependent on the availability, that is, on the formation or application of empirical concepts. I am not claiming that Kant uses the term 'experience' exclusively to refer to this basic cognitive state. Rather, I think that we should read Kant as suggesting that 'experience' comes in many shades and layers. At the (cognitively speaking) most basic layer, it *need not* involve any concepts but the schematized categories that allow for the representation (indeed: cognition⁵⁷⁰) of empirical objects *in general*. The "a priori laws which are involved in nature as such" govern a kind of synthesis of intuitions that provides us with "experience in general"; this is achieved when given intuitions are subsumed under the schematized categories, which are "fundamental concepts by which we think objects in general for appearances" (CPR, A 111). While these fundamental concepts determine an object (of human cognition) in general, empirical concepts are required to further determine and cognize the relevant object so that it can be distinguished from other empirical objects (alterations, substances):

Understanding is required for all experience and for its possibility. Its primary contribution does not consist in making the representation of objects distinct, but in making the representation of an object in general possible. (CPR, A 199/B 244).

Making possible the representation of an (empirical) object in general is the role of the categories as they are employed in the constitutive principles of the understanding: hence, thinking in terms of those principles (for instance, positing some necessitating causal ground for a perceived alteration) is constitutive of experience in general. Whether and how we can make the representation of this object distinct through empirical concept is an altogether separate issue.

My analysis of the constitutive character of the dynamical principles that involve the relational categories does not extend in a straightforward sense to the constitutive character of the mathematical principles. We already noticed two decisive differences. First, determining the concrete empirical instantiation of the categories involved in the dynamical principles requires a posteriori methods, which means that there is no guarantee that we will be able to determine what the relevant specific kinds of substances, causal powers or laws are; by contrast, the mathematical principles do not only tell us that every appearance must have some indeterminate extensive and intensive magnitude, but also provide us with an a priori method for determining the precise degree of the relevant magnitudes and hence the concrete empirical instantiation of

⁵⁷⁰ Kant says that experience, qua necessary synthesis of perceptions under the Analogies that determines an object for those perceptions (i.e., 'experience in general'), is "an empirical cognition"; see CPR, B 219; B 165.

the relevant categories. Second, the properties of objects that are determined by the mathematical categories do not suffice to characterize an object that is part of the course of nature, and correspondingly a representation of objects in terms of the mathematical categories is, as such, constitutive only of perception but not of experience.⁵⁷¹ These two features point to a crucial difficulty in Kant's conception of the mathematical principles: his tendency to conceive of these principles as rules for the construction of precise quantities or qualities is hard to square with the idea that these rules are supposed to govern, *in the first instance*, the mental act of perception.⁵⁷² It is hard to even see how Kant would want to formulate the mathematical rules for the synthesis that generates perceptions of objects; and hence it is unclear what the specifically subject-directed aspect of the mathematical principles amounts to.⁵⁷³ For these reasons, I shall focus, in the remainder of my argument, on the difference between the constitutive dynamical principles of the understanding and the regulative principles of theoretical reason.

My interpretation of the role of the dynamical principles, as being prior to and independent of the formation and correlation of empirical concepts and laws, conflicts with a number of interpretations that hold that these principles stake their claim and govern our cognitive pursuits exclusively at the level of systematic, scientific theorizing.⁵⁷⁴ I now want to

⁵⁷¹ The mathematical principles relate *only* to properties whose concepts can be constructed in pure intuition. Thus, these principles do not have *any* kind of metaphysical implications concerning which natural objects exist and what relations they bear to one another (cf. CPR, A 178/B 221). Correspondingly, the kind of representation they allow for is not even yet of a mind-independent natural object. One won't be able to represent objects in this 'weighty' sense unless one also takes into account the dynamical principles that concern the existence and dependency relations among appearances. This does not, of course, mean that the mathematical principles are irrelevant as far as higher-order thinking about natural objects is concerned: objective experience is a rule-governed combination of perceptions, and since the mathematical principles are constitutive of intuition or perceptions, it follows that they are constitutive of experience as well; see CPR, A 160/B 199 and A 172/B 214.

⁵⁷² See CPR, A 178-9/B 221 for an instance of Kant's tendency to explicate the point of the mathematical principles by appeal to an example involving the precise construction of the mathematical features of an object perception. See Anderson 2001, p. 292, Guyer 1987, pp. 195-6 and Strawson 1966, p. 31 for the claim that the mathematical principles are exclusively concerned with calculation or measurement.

⁵⁷³ A related point is that I find it rather hard to see how the subject-directed aspect of the mathematical principles could be understood in normative terms, as a rule that governs (even if only implicitly) reflective acts of thought. In what sense can an act of synthesis that generates a perceptual image be considered reflective and norm-governed? For further discussion of these difficulties, see my chapter X, especially section 4 and footnote 513.

⁵⁷⁴ As we saw, Friedman holds that the dynamical principles are primarily principles of high-level scientific cognition. Guyer 1987, p. 252 argues that the upshot of the Second Analogy is that we need to know causal laws to confirm that we have observed an objective occurrence. For conclusive criticisms, see Allison 1989 and Longuenesse 2001, pp. 337-8; I think that the main problem is that Guyer's view cannot make sense of Kant's repeated insistence that the causal principle states a condition for the objective representation of an occurrence, namely, that we must posit some causal law which we do not yet know and which we may never know. Guyer claims (*ibid.*, p. 304) that the only way in which a principle can serve as a condition for the possibility of a form of judgment is by elucidating verification conditions; but this completely ignores Kant's argument that in order for us to engage in certain forms of determinate empirical cognition (including verification), we must already be in possession of (indeterminate) representations of public empirical objects; the possibility of such representation is conditioned by the principles flowing from the schematized categories.

A special case here is Longuenesse, who claims that "reflection [and subsumption] [of appearances] under categories as concepts can occur only when empirical judgments...have been formed and systematically correlated. (...) The judgments expressing this subsumption of objects...under categories are expounded in the System of Principles of Pure Understanding" (2001, p. 244). I have discussed Longuenesse's important account at length in chapter X (see especially section 4). Here I want to notice that Longuenesse does *not* agree with

argue that my interpretation is strongly supported by what Kant says in the Introductions to the Third Critique. For there Kant emphasizes: first, that the a priori principles of the understanding relate only to nature *in general* and leave completely indeterminate the specific forms or instantiations that the general concepts that are being thought in these principles (e.g., substance, empirical lawlikeness) may take⁵⁷⁵; second, and similarly, that the Analogies relate only to experience in general rather than to a specific experience of objects as represented under empirical concepts.⁵⁷⁶ Moreover, Kant distinguishes between two kinds of unity or connectedness: namely, (a) the unity that nature exhibits *qua* being composed of objects that fall under transcendental (a priori) laws as natural objects in general, which corresponds to the unity that our experience exhibits as a systematic combination of representations in one consciousness; and (b) the unity nature exhibits *qua* being constituted by objects that fall under empirical concepts and laws, which corresponds to the unity that our experience exhibits as a systematic combination of judgments of experience and empirical concepts. Kant denies that the second kind of unity is contained in or implied by the first.⁵⁷⁷ On this basis, he holds that it is

Friedman: she rejects the notion that the Analogies are constitutive of, primarily, *scientific* experience (2009, pp. 143 ff.), and she argues in a convincing manner that endorsing a 'strong' reading of the Second Analogy that implies the 'same cause, same effect' notion of universality is compatible with reading the Second Analogy as governing a kind of experience that precedes the systematic, scientific cognition of causal laws. I find this interpretation of the Second Analogy very congenial. But it is not quite obvious to me how it coheres with her conviction (which is explicit in the above quotes) that it is only at the level of forming and systematically correlating empirical judgments that the Analogies are relevant for our cognitive orientation.

⁵⁷⁵ KU, 5: 179-80 (my emphasis): "But there are such manifold forms of nature, so many modifications, as it were, of the universal transcendental concepts of nature, left undetermined by the laws furnished by pure understanding a priori as above mentioned, and for the reason that these laws only touch the possibility of a **nature in general** (as an object of sense), that there must also be [empirical] laws in this behalf."

⁵⁷⁶ KU, 5: 184 (my emphasis): "The understanding is no doubt a priori in possession of universal laws of nature, apart from which nature would be incapable of being an object of experience at all. But over and above this it needs a certain order of nature in its particular rules which are only capable of being brought to its knowledge empirically, and which, so far as it is concerned are contingent. These rules, without which we would have no means of **advance from the general analogy of a possible experience in general to a particular** [analogy of experience], must be regarded by understanding as laws, i.e., as necessary...though it be unable to cognize or ever get an insight into their necessity." EEKU, 20: 210 (my emphasis): "But the understanding, in its transcendental legislation of nature, abstracts from all manifold of possible empirical laws; in this legislation it **only considers the conditions of the possibility of an experience in general**, according to its form."

⁵⁷⁷ KU, 5: 182-3 (my emphasis): "Now, looking at the grounds of the possibility of an experience, the first thing, of course, that meets us is something necessary—namely, the universal laws apart from which nature as such... cannot be thought. (...) For instance, understanding says: all change has its cause (universal law of nature). (...) Now for nature as such, as an object of possible experience, that law is cognized as absolutely necessary. But besides this formal time-condition, the objects of empirical cognition are determined, or, so far as we can judge a priori, are determinable, in diverse ways, so that specifically differentiated natures, over and above what they have in common as things of nature as such, are further capable of being causes in an infinite variety of ways. (...) In respect of these [empirical laws] we estimate **the unity of nature according to empirical laws**, and the possibility of **the unity of experience, as a system according to empirical laws**, to be contingent. (...) **...the universal laws of nature, while providing, certainly, for such a connection among things generically, as things of nature in general, do not do so for them specifically as such particular things of nature.**" EEKU, 20: 208-9: "We saw in the *Critique of Pure Reason* that **the whole of nature, as the epitome of all objects of experience, is a system according to transcendental laws, namely those that the understanding itself gives a priori** (for appearances, insofar as they ought to constitute, as combined in one consciousness, experience). Thus experience...insofar as it is as such objectively considered possible, must be a system of possible empirical cognition: because this is demanded by the unity of nature, according to a principle of the thoroughgoing

conceivable that the first kind of unity may take place without the second: the unity of (experience of) nature that empirical objects exhibit in general, *qua* being subsumed (and cognized) under the categories, might be coupled with a disunity of (experience of) nature that objects exhibit *qua* falling under empirical concepts.⁵⁷⁸

I conjecture that what motivates this picture is, in part, the need to temper the idealist notion of a dependency of natural objects on our faculties of knowledge. Kant was well aware of the fact that his 'revolutionary' metaphysical idea that our understanding imposes unity on nature is bound to offend our realist intuitions. It is a concession to common-sense realism to emphasize that this unity pertains only to the most basic or general empirical character of objects⁵⁷⁹: whether or not natural objects exhibit a systematic unity in virtue of their *specific* empirical constitution

combination of everything contained in this epitome of all appearances. (...) **But it does not follow from this that nature is also a system according to empirical laws** that is comprehensible to the human capacity for cognition, and that the thoroughgoing systematic connection of its appearances in one experience, therefore nature itself as a system, is possible."

⁵⁷⁸ KU, 5: 185: "For it is quite conceivable that, despite all the uniformity of the things of nature according to universal laws, without which we would not have the form of general empirical knowledge at all, the specific variety of the empirical laws of nature, with their effects, might still be so great as to make it impossible for our understanding to discover in nature an intelligible order, to divide its products into genera and species so as to avail ourselves of the principles of explanation and comprehension of one for explaining and interpreting another, and out of material coming to hand in such confusion (properly speaking only infinitely multiform and ill-adapted to our power-of apprehension) to make a consistent context of experience." EEKU, 20: 210-11: "That nature, according to its merely formal laws (whereby it is an object of experience as such), is regulated by the understanding can well be appreciated, but with respect to specific laws, its manifoldness and heterogeneity, it is free from all limitations deriving from our legislating capacity for cognition...".

⁵⁷⁹ It is a difficult question to what extent this this concession is explicit already in the First Critique. Paton 1970, p. 371 claims that Kant's account of necessary synthetic unity under the categories is already meant "to guarantee the homogeneity and regularity which we all assume to be present in nature", and that the Third Critique's emphasis on the limitations of the unity determined by the categories is an altogether novel idea. Guyer 1990 and 1992, pp. 20-4 argues along similar lines that there is a great systematic shift from First to Third Critique. But (pace Guyer 1990, p. 35) Kant emphasizes already in the First Critique (in both versions of the Deduction: CPR, B 165; A 125), that the categories enable only a *formal* unity of experience and of nature (*natura formaliter spectata*), which pertains to the lawfulness and coherence of appearances only at the most abstract and general level and falls short of determining any specific laws or concepts. Now, at one point in the Dialectic (CPR, A 653-4/B 681-2), Kant asserts that the theoretical idea of unity, which pertains to the *specific* character of appearances under empirical concepts (cf. section 4 below), is also a condition for the possibility of experience. So is Kant committed to the claim that the ideas of reason can (*qua* conditions of the possibility of experience) be given a strict transcendental deduction, just like the categories (something he explicitly denies; CPR, A 664-5/B 692-3)? I think that we can avoid this result only if we assume that there is a difference in the notions of experience that Kant uses in the Dialectic and the Analytic – a difference which anticipates the systematic difference that he draws in the Introductions to the Third Critique. Indeed, I think that there is textual support for this assumption: in the Dialectic passage that is meant to show that the theoretical idea of unity of appearances is a condition for the possibility of experience, Kant's apagogical argument proceeds from the premise that appearances are wholly chaotic "according to their content"; but it grants that appearances are unified "according to their form". This, I suggest, shows that Kant is not here revoking the claim that *only* the categories are needed to ensure that there is a formal unity of experience and nature. The result of the Deduction, as anticipated at A 90/B 123-4, is that the categories *qua* conditions of synthetic unity are not distinct from the conditions imposed upon objects by our forms of sensibility. Hence, they are conditions for the unity of experience of objects in one spatiotemporal framework. But this does not guarantee the unity of experience of objects that fit together in one system of specific empirical concepts, laws, and judgments. For *this* kind of (unity of) experience, the ideas of reason are needed (see section 4 below).

or properties is wholly independent of our cognitive constitution, i.e., of the way in which we must think about empirical objects in order to know them as such.

This picture obliges us to assign a cognitive role or function to the Analogies that can be specified in complete independence of any *successful* formation and (systematic) correlation of empirical concepts and judgments: it would be very odd, maybe even incoherent, for Kant to hold *both* (1) that the Analogies emerge, or guide us, only when we form and correlate empirical concepts or laws *and* (2) that the kind of unity or system (of both nature and experience) enabled by the Analogies is insufficient for the systematic unity of (things under) empirical concepts and laws. As I showed, Kant is strictly committed to (2). Hence we should not attribute (1) to Kant. *Metaphysically speaking*, the Analogies articulate nothing more or less than the empirical character that objects of human experience have as such and that allows them to cohere in one unified spatiotemporal framework. *Epistemically speaking*, the Analogies govern a synthesis that brings given representations under concepts of empirical objects in general and that allows them to stand together in one unified course of experience in general. This is all that the understanding, the source of the Analogies, can *by itself* contribute to our cognitive enterprise or determine with regards to the character of empirical objects. It provides us with a basic insight into the constitution that objects of our senses must exhibit, and it gives us a set of cognitive norms that we must comply with to have an equally basic kind of cognition of objects. In complying with the First and Second Analogy, that is, in positing, for any succession of states, something that persists through a change and some condition that nomologically necessitates the change, we are provided with indeterminate reference points that allow us to construct an objective temporal framework over and above the fleeting course of our sensory representations. Having constructed such a framework, we can try to determine what these reference points are, i.e., to cognize specific kinds of substances and causal powers under empirical concepts, to progress from experience in general to more specific or sophisticated form of experience.⁵⁸⁰

A further passage from the Prolegomena can be adduced to round up my account of the role of the categories in experience:

Quite another judgment [1] must therefore precede before experience can arise from perception. The given intuition must be subsumed under a concept, which determines **the form of judgment in general** in respect of the intuition, connects the empirical consciousness of the latter in a **consciousness in**

⁵⁸⁰ Providing a taxonomy of how Kant uses the concept of 'experience' is beyond the scope of my discussion here. But I want to notice that the ambiguity that I find in Kant's notion of experience is not identical to the one perceived by Guyer 1987, pp. 79-81. Guyer argues that Kant has a wholly subjective and an objective notion of experience, whereas on my interpretation both 'experience in general' and experience on the basis of specific empirical concepts are objective. Guyer argues that two passages tell in favor of his proposal. First, there is CPR, B1, where Kant says that "that knowledge of objects which we entitle experience" first arises when the understanding "combine[s] or separate[s]" given representations and thereby "works up the raw material of the sensible impressions". Here it is not clear to me what compels us to identify 'the raw material' with a kind of experience; Kant might mean only that experience qua cognition of objects first arises when this material is synthesized by the understanding. This will be the form of experience ('in general') that figures in the Analytic (this reading seems to be in play in Longuenesse 1998, pp. 26-7). The second passage is CPR, B 219, where Kant says that "in experience...perceptions come together only in accidental order". However, since (as Guyer notes) Kant has defined experience as objective only a few lines prior to this statement, one might well suppose that what Kant means to emphasize here is precisely the gap between objective experience and its subjective 'input', namely, the subjective order of perceptions from which experience (in general) arises; when he stresses that no objective order is revealed "in the perceptions themselves", we might add: 'in contrast to experience'. For more elaborate criticisms of Guyer's readings, see Engstrom 1994, p. 363.

general, and thereby procures universal validity for empirical judgments [2]; such a concept is an a priori concept of the understanding, which does nothing except merely **determine in general** the manner in which an intuition can serve for judgments [2]. (Prol, 4: 300; my square brackets and emphasis)

The judgment [1] which subsumes given intuitions under a priori concepts is epistemically prior to subsequent empirical judgments [2]: this subsumption "does nothing except" rendering the given sensible material suitable, in a most general way, for subsequent concept - and judgment-formation. Judgment [1] transforms perception into experience: it integrates the given representation into a 'consciousness in general' which is not empirical but transcendental, for in it perceptions are related to empirical objects in general – even if these objects have not yet been further determined on the basis of empirical concepts. That Kant considers the act of subsumption of intuitions under the categories a *judgment* that epistemically precedes the determination of objects under empirical concepts is a very important interpretive point that explains two textual data. First, Kant's insistence that the objective validity of the categories follows "from the...definition of a judgment in general (an act by which given representations first become cognitions of an object" (Man, 4: 475) can now be understood as alluding to the idea that the categories are required for relating given representations, through a very basic act of judgment 'in general', to empirical objects 'in general', an act through which we first recognize empirical objects as such according to their general empirical character. Second, since this cognitive act is so general and indeterminate, we can see why Kant would insist that the synthetic unity of apperception (to which given intuitions are brought by virtue of being subsumed under the categories) is epistemically prior to 'determinate' thought (CPR, B 134).

I want to conclude this section by explaining my main point in a further way that will be useful in what follows, namely, by emphasizing the significance of the constitutive principles of the understanding as necessary but insufficient criteria of *truth*. Now, there is a range of necessary but insufficient conditions for truth that I will not be concerned with, namely, what Kant refers to as the formal or merely logical, 'negative' conditions of *all* truth, as specified by what he calls general logic (cf. CPR, A 58-62/B 82-7). These conditions include principles such as the law of non-contradiction, and they do not specifically pertain to the empirical use of the understanding (they also apply to ethics, religion, etc.). By contrast, there is a part of his pure logic that Kant calls transcendental logic, and this branch of 'logic' is concerned with establishing our pure, a priori knowledge *of objects*. As such, it expounds necessary or formal conditions for *empirical truth* only (for all our theoretical knowledge of objects pertains to sensible, empirical objects). The rules it articulates yield necessary conditions for empirical truth insofar as they provide necessary conditions for relating representations to objects.⁵⁸¹ If one complies with these rules, one partakes in the activity of empirical cognition; if one violates these rules, one ceases to be engaged in the activity of empirical cognition, since in this case one's representations lose their relation to empirical objects (appearances) and their empirical content is compromised. That these subject-directed rules for combining given representations are at the same time supposed to give us a priori *cognition of objects* is the (by now familiar) upshot of Kant's idea that the rules that specify necessary condition for empirical truth are at the same time conditions of the

⁵⁸¹ CPR, A 62/B 87: "That part of transcendental logic which deals with the elements of the pure cognition yielded by understanding, and the principles without which no object can be thought, is transcendental analytic. It is a logic of truth. For no cognition can contradict it without at once losing all content, that is, all relation to any object, and therefore all truth."

character of the objects that these truths are concerned with.

This picture is exemplified in a very important passage from the Analytic:

What, then, am I to understand by the question: how the manifold may be connected in the appearance itself, which yet is nothing in itself? That which lies in the successive apprehension is here viewed as representation, while the appearance which is given to me, notwithstanding that it is nothing but the sum of these representations, is viewed as their object; and my concept, which I derive from the representations of apprehension, has to agree with it. **Since truth consists in the agreement of knowledge with the object, it will at once be seen that we can here enquire only regarding the formal conditions of empirical truth**, and that appearance, in contradistinction to the representations of apprehension, can be represented as an object distinct from them only if it stands under a rule which distinguishes it from every other apprehension and necessitates some one particular mode of connection of the manifold. The object is that in the appearance which contains the condition of this necessary rule of apprehension. (CPR, A 191/B 236; my emphasis)

The 'formal' condition of empirical truth that Kant mentions here is that we must draw a distinction between the subjective course of our private mental representations and an object that is something distinct from these representations; among other things, this requires drawing a distinction between the ever successive course of our perceptions and objective successions, i.e., events that happen in the (empirically, not transcendently speaking) mind-independent world. The Second Analogy (qua rule of synthesis) states that the representation of an objective change requires positing a condition from which the change follows necessarily and invariably. Complying with this rule is a minimally necessary condition for making a true judgment about the causes of events and the laws that govern the causality of these causes, because it allows one to represent objective changes (as distinct from the subjective course of apprehension) and an objective order of time. Compliance with the rule is hence constitutive of experience of alterations as such. At the same time, the rule itself is constitutive of the character of (successive) alteration as such. *Nothing else* follows from compliance with the rule for the character of our experience, or from the rule itself for the character of the succession: the rule implies nothing about the specific causes of the succession or about the specific laws governing the causality of these causes, and (correspondingly) compliance with the rule does not put us in a position to cognize the specific causes of events or the specific laws governing their causality. Such cognition requires going beyond the formal, necessary conditions for empirical truth.

The very special sense in which the principles established by transcendental logic are 'constitutive' cannot be emphasized enough. The constitutive character of these principles hinges, crucially, on the notion that they are determinative of empirical objects (in general); and correspondingly, subjecting our thoughts to these principles puts us in a position to *determine an object for* our given intuitions and to get into a cognitively significant mental state, experience in general, that allows us to represent and thereby to know things according to their general empirical character. Kant's notion of constitutive rules here must sharply be separated from a more contemporary sense in which rules may be said to be constitutive. The rules established by general logic, such as the principle of non-contradiction, are *not* constitutive in Kant's special sense of the term, precisely because these rules do not determine the character of objects and complying with them does not by itself allow us to know objects according to their character.

3: Regulative epistemic norms: the negative function

So far I explained the constitutive character of the principles of the understanding by appeal to their metaphysical and their epistemological implications. I now want to try to explain the merely regulative of the principles of theoretical reason in a similar fashion. I shall argue that these principles have a negative metaphysical and a positive epistemological function.

Kant distinguishes between two separate aspects, or functions, of regulative principles:

For these principles serve partly to restrain the officious pretensions of understanding, which, presuming on its ability to supply a priori the conditions of the possibility of all things which it is capable of knowing, behaves as if it had thus determined these bounds as those of the possibility of all things generally, and partly also to lead understanding, in its study of nature, according to a principle of completeness, unattainable as this remains for it, and so to promote the ultimate aim of all knowledge. (KU, 5: 167-8)

I want to begin by discussing the negative point of regulative principles, i.e., the constraint they put on "the officious pretensions of understanding". These pretensions result when the understanding gives in to the tendency (as Kant puts it elsewhere) "to make principles of possible experience conditions of the possibility of things in general" (CPR, A 781/B 809), that is, its tendency to assume that the objects of cognition (that necessarily fall under the constitutive principles of the understanding) are things in themselves.

Kant never really tells us how, precisely, the regulative principles serve to constrain this pretension. i.e., how they counteract our tendency to be or become transcendental realists. I think that to understand his point we must have a close look at those sections of the First Critique that precede the solutions to the antinomies. Kant thinks that the general erroneous notion underlying all the internal contradictions (antinomies) of reason is the idea that if an empirical object exists (or: if an empirical event occurs), then the entire series of its metaphysical and explanatory 'conditions' also exists independently of any reference to the subject of cognition (CPR, A 497/B 525). This idea would be correct if empirical objects of cognition were things in themselves⁵⁸², and this is an assumption toward which the understanding in its empirical employment is drawn because it is insensitive to the distinction between things in themselves and appearances:

...if the conditioned as well as its condition are things in themselves, then upon the former being given, the regress to the latter is not only set as a task, but therewith already really given. And since this holds of all members of the series, the complete series of the conditions, and therefore the unconditioned, is given therewith, or rather is presupposed in view of the fact that the conditioned, which is only possible through the complete series, is given. The synthesis of the conditioned with its condition is here a synthesis of the mere understanding, which represents things as they are, without considering whether and how we can obtain knowledge of them. (CPR, A 498/B 526-7) ...the understanding, occupied merely with its empirical employment, and not reflecting upon the sources of its own knowledge, may indeed get along quite satisfactorily, [but] there is yet one task to which it is not equal, that, namely, of determining the limits of its employment, and of knowing what it is that may lie within and what it is that lies without its own proper sphere. (CPR, A 238/B 297)

Someone who collapses the distinction between appearances and things in themselves (a realist) will assume, when faced with a given (observed) object, that the entire series of parts that compose this object, or the entire series of conditions on which this object depends (for its

⁵⁸² As Allison 2004, p. 385 puts it: if transcendental realism is true, then "the totality of conditions...exists in timeless manner as a thing in itself".

existence and its state at a given time), is also 'given', i.e., that this series exists as a complete, mind-independent totality. The exact shape of this realist assumption is determined by the thinker's antecedent, empiricist or rationalist commitments. For the rationalist, the series of parts or conditions is *finite* and culminates in some unconditioned member or final (simple) part; the empiricist treats the series as *infinite*, so that it is absolutely impossible for there to be a final (simple) part or an unconditioned condition. Against both kinds of realist pictures, Kant emphasizes that we do not know objects except insofar as they can appear before our senses. Since the *entire* series of conditions for an observed, conditioned object (or event) cannot ever be an object of sensible intuition (neither qua unconditioned condition nor qua infinite series of conditions), we cannot claim to know the character of this series: both the rationalist and the empiricist affirm claims about the totality of conditions for a given conditioned object that go beyond the limits of human knowledge. That is, we cannot *objectively* represent the series of conditions (for some conditioned object) as a complete, already 'given' totality that exists in itself, independently of a reference to the cognizing subject. This does not mean that the members of this series (appearances, i.e., empirical objects or events) depend for their existence on *actually being* cognized. Rather, it means that these members owe their reality to their status as objects of a *possible* experience, as something that we (considered as a thinking species) could encounter in the course of our human experience.⁵⁸³

This line of thought transforms the series of conditions for a given conditioned object from an actual complete totality into a chain of possible experience that must be *realized as actual* on the basis of the cognizing subject's epistemic activity. Likewise, it transforms a metaphysical claim about what the totality of a series of conditions is like (finite or infinite) into a principle "which serves as a rule, postulating what we ought to do in the regress", namely, to "undertake and to carry on, the regress in the series of conditions of any given conditioned" (CPR, A 508/B 536), ad infinitum or indefinitum (CPR, A 523-4/B 551-2), in order to gain an ever more complete understanding of the course of nature. This principle is an epistemic norm addressed to someone occupying the empirical standpoint. In this respect, it is like the principles of the understanding. But unlike these principles, it has no metaphysical implications concerning the objects of cognition: it does not imply, for instance, that for every given event there exists an infinite (or a finite) series of antecedent conditions. Since it lacks those metaphysical implications, it is not a constitutive but a merely regulative epistemic norm.⁵⁸⁴

These considerations are necessary for understanding Kant's idea that the regulative principles of theoretical reason serve the negative function to constrain the officious pretensions

⁵⁸³ CPR, A 495/B 523: "Thus we can say that the real things of past time are given in the transcendental object of experience; but they are objects for me and real in past time only in so far as I represent to myself...that a regressive series of possible perceptions in accordance with empirical laws, in a word, that the course of the world, conducts us to a past time-series as condition of the present time -- a series which, however, can be represented as actual not in itself but only in the connection of a possible experience. Accordingly, all events which have taken place in the immense periods that have preceded my own existence mean really nothing but the possibility of extending the chain of experience from the present perception back to the conditions which determine this perception in respect of time."

⁵⁸⁴ CPR, A 509/B 537: "Thus it is a principle of reason which serves as a rule, postulating what we ought to do in the regress, but not anticipating what is present in the object as it is in itself, prior to all regress. Accordingly I entitle it a regulative principle of reason, to distinguish it from the principle of the absolute totality of the series of conditions, viewed as actually present in the object (that is, in the appearances), which would be a constitutive cosmological principle."

of the understanding. The best way to see what I take to be Kant's point here is by imagining a realist version of the normative principle that states that one ought to 'carry on, the regress in the series of conditions of any given conditioned'. Surely, the realist must accept that this is indeed a norm to which everyone who aims at understanding the course of nature (and hence at theoretical knowledge) is committed. Now, the realist holds that objects of cognition are completely independent of the subjects of cognition and of their place in a possible experience. And this implies that every member of the series, and hence the entire series, must be regarded as an 'entity', an existing whole of reality. But then the realist's version of the norm that prescribes that one aims at a more and more complete grasp of the series of natural conditions inevitably raises a question concerning the status of the series qua whole of reality (or totality of conditions): is it a finite or an infinite whole? If the realist advancing this norm has rationalist sympathies (advocating the theses of the Antinomies), the independently existing series of conditions will be represented as finite, culminating in some empirically *unconditioned* member. But this requires positing an object within nature that does not satisfy the conditions for the possibility of experience (such as being in space and time); and this destroys our understanding of nature. The empiricist, by contrast, thinks that *whatever* object exists or whatever event occurs must fit into the spatiotemporal, deterministic order of nature, and hence claims that the series of conditions must be represented as infinite: every event is causally, temporally conditioned by other events. This, however, saddles reason with irresolvable internal conflicts.⁵⁸⁵

Hence, if the norm to carry on in the empirical regress treats the series of conditions that are to be met with in this regress as a complete (either finite or infinite) totality, severe problems arise. A 'critical' version of this norm will emphasize that we cannot represent the series of empirical conditions as a complete totality that already exists (that is 'given') independently of its relation to our cognitive faculties. This prohibition serves to undercut (i.e., exposes as invalid) the inference from the fact that some conditioned object is given to the notion that, therefore, the totality of its conditions is also given.⁵⁸⁶ Hence, someone whose epistemic activity is governed

⁵⁸⁵ Kant is clear about the fact that dogmatic empiricism annihilate the possibility of practical freedom and hence of morality (CPR, A 472/B 500), and thus the empiricist position throws reason into an insoluble conflict concerning its theoretical and practical use. But it is important to see that Kant thinks that dogmatic empiricism also throws the *theoretical* employment of reason into conflict. He states (CPR, A 475/B 503) that according to such empiricism the completion of a 'finished structure' of a system of empirical cognitions is "wholly impossible", because it deems strictly (metaphysically) impossible the notion of a first beginning or of an unconditioned condition; and this conflicts with the architectonic interest of reason in its pursuit of complete systematicity (i.e., it does not allow us to treat a complete system of empirical cognitions *even as an ideal* to which we can asymptotically approximate). This may not seem all that compelling (especially since Kant himself has to admit that a complete system of empirical concepts remains *only ever* an ideal for us). But there is a more specific version of this general point which may be more forceful: the empiricist's commitment to an infinite amount of complexity of conditions, causes and parts in nature is connected to a commitment to the notion that there is an infinite manifold of forms to be found in nature. Given her realist outlook, she will construe the notion that nature is infinitely heterogeneous as a claim about the inner constitution of nature; as such, however, it is incompatible with the further notion that nature is ultimately *homogeneous*, and rejecting that latter notion is, Kant argues, detrimental to the systematic pursuit of empirical knowledge (cf. CPR, A 654-5/B 682-3 ff.; and see my chapter IX, section 4). In addition, Kant can also point out that empiricism negates the possibility of establishing the validity of principles such as, 'every event follows necessarily from a preceding cause', and hence renders impossible the very commitment to the kind of universal naturalistic determinism that is definitive of empiricist realism. This is the point of Kant's claim (CPR, A 543/B 571) that if transcendental realism is assumed, *neither* nature (*qua* series of alterations governed by deterministic laws) *nor* freedom will remain.

⁵⁸⁶ "If, however, what we are dealing with are appearances – that, as mere representations cannot be given save in so

by the critical norm does not need to choose between representing the series of conditions as something that already exists in either finite or infinite form, that is, between either positing an unconditioned causality within nature or dogmatically declaring that such an unconditioned causality is metaphysically impossible. She thus avoids the internal contradictions that someone who is governed by the realist version of the norm inevitably slides into.

So, I interpret Kant as holding that the regulative normative principle that prescribes continuing in the regress to further and further empirical conditions tempers the officious, realist pretensions of the understanding as follows: it includes, in its content, a reminder (or rather an *alert*) that the status of the conditions for any given conditioned object is nothing but that of objects of a possible experience. It thereby emphasizes that we must not represent the series of conditions in 'objective' terms as a complete, given totality, but in 'subjective' terms as something to be achieved, as a cognitive task that we are faced with:

Since no maximum of the series of conditions in a sensible world, regarded as a thing in itself, is given through the cosmological principle of totality, but can only be set as a task that calls for regress in the series of conditions, the principle of pure reason has to be amended in these terms; and it then preserves its validity, not indeed as the axiom that we think the totality as actually in the object, but as a problem for the understanding, and therefore for the subject, leading it to undertake and to carry on, in accordance with the completeness prescribed by the idea, the regress in the series of conditions of any given conditioned. (CPR, A 508/B 536)

On the basis of this interpretation, we can give substance to Kant's claim (GMS, 4: 452) that the spontaneity of reason has a higher authority than that of the understanding insofar as the ideas of reason 'prescribe' ("vorzeichnen") limits to the understanding. The ideas of theoretical reason are epistemic norms that deem the understanding's identification of its objects of experience with things in themselves *epistemically impermissible*. They do so by reminding the understanding of the fact that its concepts and rules pertain only to appearances or, what amounts to the same, by alerting it to the fact that the objects of our cognition owe their status as *objects* solely to their place within a possible experience. So theoretical reason has normative authority over the understanding, an authority which derives from its recognition of the status of objects of human cognition and (hence) of the limits of human cognition. Through exercising this authority, reason warns us not think about reality in a subtly but gravely distorted fashion.

4: Regulative epistemic norms: the positive function

We saw that the two aspects or functions of the constitutive principles of the understanding are closely interrelated: they are, at the same time, (metaphysically speaking) conditions on objects of experience and (epistemically speaking) on experience of objects. There can be no such close link between the two aspects of regulative principles: since their metaphysical upshot is negative and reduces to an admonition *not to* treat objects as things in themselves, their positive role in our epistemic pursuits must be specified independently of this admonition.

To get a grip on what this positive role is, we need to remember that the constitutive rules of the understanding govern only basic cognitive acts of experience of empirical objects in

far as I attain knowledge of them, or rather attain them in themselves, for they are nothing but empirical modes of knowledge – I cannot say...that if the conditioned is given, all its conditions (as appearances) are likewise given, and therefore cannot in any way infer the absolute totality of the series of its conditions. (...) It does not, therefore, follow, that if the conditioned, in the [field of] appearance, is given, the synthesis which constitutes its empirical condition is given therewith and is presupposed" (CPR, A 498-9/B 527).

general. They cannot by themselves establish, or guide us in establishing, those rules that pertain to the specific character that distinguishes an empirical objects among others, namely, *empirical concepts*. In order to form valid empirical rules or concepts, our epistemic agency stands in need of governance by norms other than those provided by the understanding. This is how the need for regulative epistemic norms arises: theoretical reason is there "to assist the understanding by means of ideas, in those cases in which the understanding cannot by itself establish rules" (CPR, A 648/B 676). Kant thinks that these rules are concerned with empirical concept *formation*: namely, with the ascent from given representations (intuitions or concepts) to more general representations that allow us to (initially or further) determine, classify and hence cognize the initial representations (and their objects).⁵⁸⁷ Kant argues that the conscious effort to devise and revise our scheme of empirical concepts stands in need of principles that prescribe that in our attempts to find empirical concepts and laws, we should make certain assumptions about the way nature is ordered and structured. These principles have the status of regulative epistemic norms.

I suggest that in order to understand how these norms govern epistemic activity, we must begin with the very general principle that we have examined in the preceding section, which prescribes that we ought to "carry on, the regress in the series of conditions of any given conditioned". This principle (on my view) gives us the *common general core* of the *various specific* regulative principles that Kant discusses in the section entitled, 'Of the regulative use of ideas of pure reason' (CPR, A 642/B 670 ff.). The general, common core of all regulative epistemic principles consists in the prescription – directed toward anyone occupying the empirical standpoint – to gain a more and more thorough understanding of the course of nature by advancing further and further in the regress from a given observed object (state, event) to its conditions. Kant reminds us of this general core when he says that the common point of all regulative principles is "to direct the understanding beyond every given experience...and thereby to secure its greatest possible extension" (CPR, A 644-5/B 672-3). This direction is as general as it is empty: it does not specify *how* we ought to undertake the regress from conditioned to conditions, that is, how we might achieve the greatest possible extension of the understanding.

The need for more concrete epistemic prescription is answered by Kant's introduction of the "regulative law of systematic unity" that "prescribes that we ought to study nature *as if* systematic and purposive unity, combined with the greatest possible manifoldness, were everywhere to be met with, *in infinitum*" (CPR, A 701/B 729). This regulative law is itself a rather general summary version of three more determinate norms: first, the norm of *homogeneity*, which prescribes that we should always look for the greatest possible unity of empirical concepts and laws, and that we should attempt to trace divergent empirical powers and laws to as few fundamental powers and laws as possible; second, the norm of *specification*, which prescribes that we should always look for the greatest possible diversity in the laws and concepts that may fall under a common genus; and third, the norm of *affinity*, which prescribes that we should always, in our cognitive ascent from species to genera as well as in our cognitive descent from

⁵⁸⁷ The concern here is with the reflective use of our power of judgment. This is to be contrasted with the subsumptive use of the power of judgment, which concerns the application of given general representations to less general representations (concepts or intuitions). See KU, 5: 179-81 for a clear statement of this distinction. In the First Critique, a very similar distinction is drawn between the apodictic and the hypothetical use of theoretical reason (CPR, A 647/B 675). For the purposes of this chapter, that is, of drawing a contrast between constitutive and epistemic norms and principles, I will assimilate the principles of theoretical reason and of reflective judgment as instances of (merely) regulative principles.

genera to species, look for continuity in the various empirical concepts and laws that we have at our disposal. These three epistemic norms are intended to jointly guide us in our pursuit of the cognitive ideal of systematic unity of our various empirical cognitions (CPR, A 659/B 687).

A full examination of the justification of these normative principles is well beyond the scope of this chapter. But we can say at least this: the justifiability of our commitment to the specific norms of homogeneity, specification, and affinity depends on the notion that abiding by these norms ensures that our disparate empirical cognitions are systematically connected with one another. This seems like a plausible idea, given the content of the three norms. But even if this idea seem reasonable, Kant's line of thought here raises the further question of why we ought to be committed to the systematic unity of our cognitive pursuits. Kant obviously thinks that attempting to unify our system of cognitions – trying to determine how things hang together evidentially and causally – is the only secure and reliable means of achieving and extending our empirical cognition. But why think that striving toward systematicity among one's separate cognitions is a precondition for reliably achieving and extending our knowledge about the empirical world? To this question, Kant does have at least one interesting response, and this brings us back to a point that emerged in section 2 of this chapter.

We saw that Kant holds that the constitutive principles of the understanding articulate necessary but insufficient conditions for empirical truth. We can now see that he thinks that the regulative principles of theoretical reason add sufficiency to our criteria for empirical truth (which, in turn, shows that they cannot be established – or verified as true – on the basis of empirical observation):

For with what right can reason...call upon us to treat the multiplicity of powers exhibited in nature as simply a disguised unity and to derive this unity, so far as may be possible, from a fundamental power -- how can reason do this, if it be free to admit...that such systematic unity of derivation may not be in conformity with nature? Reason would then run counter to its own vocation, proposing as its aim an idea quite inconsistent with the constitution of nature. Nor can we say that reason, while proceeding in accordance with its own principles, has arrived at knowledge of this unity through observation of the accidental constitution of nature. The law of reason which requires us to seek for this unity, is a necessary law, since without it we should have...no coherent employment of the understanding, **and in the absence of this no sufficient criterion of empirical truth.** In order, therefore, to secure an empirical criterion we have no option save to presuppose the systematic unity of nature as objectively valid and necessary. (CPR, A 651/B 679; my emphasis)

Unless we takes ourselves to be constrained, in our formation of empirical concepts, by the assumption that things in nature constitute a unified system, we cannot take ourselves to be justified in thinking that some rather than other empirical concepts apply to reality, and that some rather than other empirical judgments (in which these concepts are applied to reality) are true. That is, unless we strive for coherence in our various disparate empirical findings, we cannot regard these findings as revealing to us what the empirical world is really like.

The idea that the coherence and unity of a system of empirical concepts or judgments yields a 'sufficient' criterion for empirical truth raises enormous philosophical questions and difficulties. Some of these are familiar from contemporary discussions of coherentism, such as John McDowell's worry that Donald Davidson's coherentism about empirical knowledge would imply that our empirical cognition is a mere 'frictionless spinning in the void', unconstrained by

sensible data.⁵⁸⁸ Applied to Kant's position, the worry can be taken to raise the point that if it is sufficient for the objective reality of an empirical concept, or for the truth of a judgment employing this concept, that it has a proper place within a systematically connected and unified web of concepts or judgments, there is no need for an additional constraint on concept or judgment formation by sensible, worldly input. As Hannah Ginsborg aptly put the worry to me, it seems that empirical truth *either* lies in agreement of our judgments with given sensible intuition *or* in their systematic connectedness. One cannot, it seems, have it both ways.

Adequately addressing this worry would require a study of its own, but I want to sketch how Kant might respond to it. One point worth emphasizing is that Kant's regulative principles that prescribe the pursuit of systematic unity provide sufficient criteria for empirical truth *only under* the presumption that the constitutive norms *qua* necessary conditions of empirical truth are already satisfied; and the concepts employed in these constitutive principles – the categories – strictly depend for *their* objective reality on sensible conditions (the schemata; cf. my chapter IX, section 4). Complying in one's epistemic activity with the constitutive norms of the understanding is a necessary condition for partaking in the activity of forming concepts and judgments about empirical objects; a proper (rule-governed) synthesis of *given sensible* intuitions must precede empirical concept formation.

But the constraint that the schematized categories impose on our epistemic activity pertains merely to the *most general* conditions of human sensibility. This does not defuse the worry that the appeal to systematic unity replaces a constraint on empirical concepts and judgments that comes from the *specific* content of what is given through sensibility. I think that Kant's claim that we need the epistemic guidance of the ideal of systematic unity rests on his conviction that sensible data and constitutive principles are *jointly insufficient* for empirical concept formation. To see why, consider our search for specific kinds of causal laws and powers. The category of causality informs us that the laws governing the causality of empirical causes must be *necessitating* and *invariable*, but neither empirical necessity nor empirical lawlikeness can be directly read off from our perceptual data: empirical necessity cannot be sensibly intuited⁵⁸⁹, and our perceptions show us only that one type of event regularly follows another, which does not allow us to separate lawlike from accidental correlations, or, within the former class, surface phenomena from those effects that are imminently produced by (and hence indicate) the cause. Hence, our cognitive efforts need criteria for determining the specific instantiations of empirical necessity and lawlikeness that cannot be derived either from the pure concepts of the understanding or from the content of the sensible given.

I want to sketch two different ways in which the notions grouped under the ideal of systematic unity (homogeneity, specificity, affinity) may function as such criteria. The first point is that these notions give our epistemic activity a *sense of direction*. Suppose we are interested in the properties, and specifically the causal powers, of a newly discovered fluid substance s1. We can only know the causal powers of substances by their effects, so in trying to get a grip on the causal powers of s1 we must observe its effects. Now, no experiment here would bear any cognitive fruits unless we knew what we were looking for: we could not design an experimental setting unless we had a prior sense of the what the effects of s1 might be, and the observation of

⁵⁸⁸ See McDowell 1994.

⁵⁸⁹ CPR, A 137-8/B 176-7: "...no one will say that a category, such as that of causality, can be intuited through sense".

some event that regularly follows (say) the impact of s1 on a surface would be cognitively empty for us unless we had a prior sense of what this observation tells for or against. Here the assumption that substances and causal powers in nature are *homogeneous* provides guidance: it implies that the causal powers of s1 can be informatively related to the causal powers of other substances that we already know on the basis of their effects, and it directs us to devise experiments to test for the manner in which this homogeneity is instantiated.⁵⁹⁰ We might start off with the hypothesis that the causal powers of s1 are sufficiently like the causal powers of other, at least superficially similar substances; on this basis we look (i.e., test) for ways in which the effects of s1 are like the effects of other fluids. This, however, leaves a decisive role for perception. For it might turn out that our predictive expectations are frustrated by the experiment and that the perceivable effects of s1 are altogether unlike those of apparently similar substances – in which case we have to judge that the similarity between s1 and other fluid substances is only superficial. In such a situation, we are constrained to continue our search for homogeneity in a different way, namely, say, to look for less obvious ways in which the causal structure of various kinds of fluids coheres, or to find homogeneity between s1 and substances that appear, at first glance, to be very different, up to the point where our very (initial) concept of s1 changes:

And indeed what useful purpose could be served by defining an empirical concept, such, for instance, as that of water? When we speak of water and its properties, we do not stop short at what is thought in the word, water, but proceed to experiments. (CPR, A 728/B 756)

The aim of forming or revising empirical concepts is not furthered by armchair, a priori conceptual analysis but by conducting experiments under the direction of the regulative a priori ideas that compose the ideal of systematic unity, i.e., by assuming that there is (e.g.) homogeneity in nature and by testing for the concrete ways in which this homogeneity is instantiated. Obviously, the ideal of systematic unity here does not license the invention of concepts that beautifully enhance the internal harmony of our system of cognition. Rather, we should understand Kant as holding that the individual norms implied by this ideal are (*unlike* the rules of the understanding that specify necessary conditions for experience) *sufficient for providing our quest for empirical truth with a goal or direction*, that is, with an idea of *what to look for* when we attend to what experience reveals about nature.

This gives us a rather modest but nonetheless, I think, informative construal of the claim that the ideal of systematic unity adds sufficiency to our criteria for empirical truth. I now want to suggest that Kant understands this claim also in a second, related but somewhat more robust way. Here the point is that a given observation has evidential significance for us *only insofar as* it can be related to our overall web of empirical concepts and laws. This is to say: a sensible datum may confirm or disconfirm a certain statement, or tell for or against a certain conceptualization, only if we consider it in the light of our existing cognitive framework and hence assume that the disparate natural phenomena we encounter in experience are systematically related to one another. Consider, for instance, a perception (artificially) stripped from all content but its categorial and sensational features; say, that of a moving, roundly shaped black body. Such a perception could not support the formation of any cognitively significant empirical concept unless the subject draws on her scheme of empirical concepts (or laws) and

⁵⁹⁰ We also assume that s1, as a *unique* type of stuff, has unique causal properties (norm of specification), and on this basis devise experiments to test for the manner in which this *specificity* is instantiated.

considers the object of the perception as (say) a type of animal, insect, beetle, etc. Unless the perception were reflected under concepts drawn from our systematic understanding of nature, and unless we took the observed phenomenon to be causally and evidentially connected to previously observed and conceptualized phenomena, the perception could not exert a rational constraint on our empirical thought. The supposition that the perception reveals a new species, and hence supports the introduction of a novel empirical concept, calls for further observation which is directed toward discovering familiarities and differences between the (putative) new species and known species, and which is hence guided by the assumptions that natural organisms are systematically connected through certain common properties but also endlessly diverse.

Consider, next, the observation that some event A regularly precedes event B. This observation is by itself insufficient to support the positing of a lawlike generalization; the notion that we have discovered a genuine law rather than a mere correlation must be supported on the basis of holistically oriented empirical reasoning.⁵⁹¹ Such reasoning considers, on the one hand, whether the putative causal law informatively relates to (i.e., shares common features with) those laws and powers that *are* an established part of our cognitive orientation; and, on the other hand, whether the putative causal law has an explanatory value that is not made redundant by those known laws (and powers). Our entitlement to the notion that the observed regularity has (unobservable) modal force is justified (first) by our unassailable knowledge that nature is inherently imbued with such modal force and (second) by the fact that the regularity we are considering fits into and enhances our system of empirical knowledge. By virtue of the Second Analogy, we must posit some universal law for every alteration; the observed regularity can be deemed suitable for filling this cognitive gap if it is entailed/confirmed by and entails/confirms other parts of our system of empirical understanding.⁵⁹²

We could also consider cases in which an experience is taken to *disconfirm* some empirical judgment or law; here the point of Kant's holism is not, I think, that we can 'choose' to discard such a 'recalcitrant' experience in the light of our overall commitments (this would be a Quinean holism⁵⁹³), but rather that the sensible impression acquires its disconfirming force only within the context of a hierarchically structured system of cognition: it is only under the presumption that natural phenomena are causally and evidentially related, with certain laws and powers being causally/explanatorily fundamental and other (pseudo-)properties at the periphery, that we take a carefully designed scientific experiment to falsify naïve views about the paths of the movements of the planets but do not (or rather *should* not) take our visual perception of the colors of things to disconfirm the scientific view that objects are not intrinsically colored.

I have suggested two ways of understanding Kant's claim that systematic unity is needed to add sufficiency to our criteria for empirical truth. Neither of those interpretations implies that systematic unity somehow competes with sensible input for the role of putting a constraint on empirical concept formation. Rather, Kant's point is that sensible data, even when

⁵⁹¹ Armstrong 1983, p. 55 suggests that the belief that it is a law that all Fs are G is made rational by the fact that all observed Fs are G and "further observational circumstances". He does not indicate what those circumstances would be, and it is hard to see how observation could support the existence of law-likeness apart from informing us about regularities.

⁵⁹² Kant seems to suggest that considerations of systematicity are needed to justify our entitlement to the idea that certain observed regularities (rather than others, we can add) have the status of necessary laws at KU, 5: 179-80. See Guyer 2003, pp. 285-6 and Rush 2000, p. 842.

⁵⁹³ For a more direct comparison of Kant's and Quine's holism, see chapter IX, footnote 471.

already brought under the categories and hence *qua* experience *in general*, must be considered in the light of the ideas of theoretical reason in order to function as empirical evidence. Without the direction provided for our quest for empirical truth by theoretical reason, sensible intuitions would be *evidentially blind*.⁵⁹⁴ This does not render the specific content of sensible intuitions evidentially obsolete, and it does not prevent empirical observation from putting a genuine constraint on our concept and judgment formation. Nature must be our witness in trying to determine the way the empirical world is, but in order to learn from this witness, theoretical reason, as its judge, must put the right questions before it:

...reason has insight only into that which it produces after a plan of its own, and that it must not allow itself to be kept, as it were, in nature's leading-strings, but must itself show the way with principles of judgment based upon fixed laws, constraining nature to give answers to questions of reason's own determining. Accidental observations, made in obedience to no previously thought-out plan, can never be made to yield a necessary law.... Reason, holding in one hand its principles, according to which alone concordant appearances can be admitted as equivalent to laws, and in the other hand the experiment which it has devised in conformity with these principles, must approach nature in order to be taught by it. It must not, however, do so in the character of a pupil who listens to everything that the teacher chooses to say, but of an appointed judge who compels the witness to answer questions which he has himself formulated. (...) ...while reason must seek in nature, not fictitiously ascribe to it, whatever as not being knowable through reason's own resources has to be learnt, if learnt at all, only from nature, it must adopt as its guide, in so seeking, that which it has itself put into nature. (CPR, B XII-XIV)

It is no accident that Kant in this passage does not mention the understanding. The "principles of judgments based upon fixed laws" he appeals to here are the ideas of theoretical reason which provide our quest for empirical truth with sufficient orientation and focus.

⁵⁹⁴ This blindness differs from the 'blindness without concepts' that Kant ascribes to intuitions at the beginning of the *Analytic*. The 'evidential blindness' I have in mind cannot (unlike the blindness Kant is concerned with in the *Analytic*) be cured by the subsumption of intuitions under the categories. I want to briefly sketch how my interpretation relates to some important debates in recent Kant scholarship, namely the debates over whether intuitions in Kant have conceptual content, whether (in particular) the categories are already involved in perception, and how we get from intuitions to empirical concepts. I agree with Ginsborg 2008 (against Hannah 2001) that in Kant perceptions are already conceptually structured; however, I think that the *only* concepts that necessarily structure perceptual representations *as such* are the mathematical categories; the application of the dynamical categories marks the transgression from perceptual representation to experience in general (see chapter X, section 4). On my view, neither perception nor experience in general requires the application of empirical concepts, beyond very indeterminate concepts such that of alteration. This highly schematic picture might provide a partial response to the important challenge that Ginsborg has recently posed to the Kantian picture. In her 2006 she points out that it must be shown, on Kant's behalf, that what is given through the senses is (1) determinate enough to provide us with rational resources for the formation of empirical concepts but (2) not so overly determinate that Kant ends up assimilating sensory, passive experience to empirical judgment, and hence begging the question against his empiricist predecessors. With regards to (2), I suggest that Kant indeed holds that experience is not a wholly passive matter, but in doing so he does not beg the question against the empiricist because the active (spontaneous) element in *experience in general*, i.e., the subsumption of appearances under categories, provides for the kind of spatiotemporal ordering that empiricists all too often simply take for granted. Hume, for instance, concedes that space and time, as ways in which impressions are ordered, cannot themselves be derived from impressions, but he has no positive account for the origin of ideas of space and time (see his *Treatise of Human Nature*, book I, part II). With regards to (1), I think that Kant's point is that intuitions can put a rational constraint on empirical concept formation only if they are reflected under some already *available* (pure and empirical) concepts and only if we are guided by the holistic ideas of reason. This is, obviously, only a first sketch toward a solution to the problems raised by Ginsborg.

4: The contrast between constitutive and regulative epistemic norms

We are now in a position to state with some precision the difference between constitutive and regulative epistemic norms. The first difference pertains to the way in which these two types of norms relate to *objects of cognition*. Both constitutive and regulative epistemic norms require us to make certain very general and indeterminate assumptions about nature: for instance, in the case of constitutive norms, the assumption that every event follows necessarily from some preceding event according to an invariable rule; or, in the case of regulative norms, the assumption that natural phenomena are related to one another in terms of a systematic unity.⁵⁹⁵ The decisive difference concerns the *epistemic status* and the *metaphysical implications* of these respective assumptions. In the case of constitutive norms, there is a metaphysical import: we *know* that nature really must be the way in which these norms urge us to think about nature (e.g., we know that every alteration does follow necessarily from some preceding event according to an invariable rule). In the case of regulative norms, there is no metaphysical import: we do not know that nature really must be the way in which these norms urge us to think about nature, i.e., that natural phenomena *are* related to one another in terms of homogeneity, specification, and affinity. I said that constitutive normative principles are 'ontologically transparent' to the way empirical objects necessarily must be; there is no such ontological transparency in the case of regulative normative principles. As Kant puts it (CPR, A 666/B 694), regulative principles are 'subjective' in the sense that they derive their legitimacy solely from the speculative demands of theoretical reason, rather than from any insight into the constitution of objects of experience.⁵⁹⁶

⁵⁹⁵ Kant also thinks that full compliance with regulative norms requires that one hypothetically and heuristically posits the existence of certain objects (soul, world, and God; see CPR, A 669/B 697 ff.). For instance, he thinks that assuming the systematic unity of natural phenomena requires assuming that there is a highest being who arranged all natural phenomena in a systematically interconnected manner that answers to our cognitive purposes. This point raises interpretive and philosophical challenges that I cannot deal with here.

⁵⁹⁶ There is considerable debate about the 'subjective' or 'objective' status of the principles of reason. The extreme poles are occupied by (a) realist interpretations that take these principles to yield objective descriptions of nature (e.g., Abela 2002; Wartenberg 1992) and (b) methodological interpretations according to which these principles derive their legitimacy solely from the heuristic value they have for our cognitive purposes (e.g., Guyer 1990, with regard to the First Critique; Philip Kitcher 1986; Vaihinger 1949). With one caveat, I side here with proponents of (b): (a) faces too many textual obstacles, and it cannot make sense of Kant's repeated insistence that the principles of reason must not be considered constitutive, or that an objective construal of these principles could lead into contradiction (CPR, A 666/B 694). The caveat is that proponents of (b) tend to underplay the extent to which Kant thinks that reason urges us to adopt *firm factual beliefs about the way nature really is*. A mere heuristic assumption is in principle replaceable and dispensable, and it requires no conviction – by contrast, for Kant we must accept that empirical phenomena *are* unified, and we must hold on to this belief whatever experience may reveal to us. This point comes to light in Kant's distinction – which underlies the objectivist interpretation (a) – between a merely logical and a transcendental version of the principles of reason (cf. CPR, A 656/B 684); the transcendental version states that we must presuppose (e.g.) the idea of systematic unity of nature "as objectively valid and necessary" (CPR, A 651/B 680). I suggest the following interpretation: for Kant, the principles of reason are not mere rules of inference that apply to any set of concepts regardless of the content of those concepts. Rather, they are transcendental in that they contain a priori assumptions about objects of our knowledge: they oblige us to assume that things in nature are systematically unified and endlessly specific, and that we must always reason, experiment and judge on the basis of this assumption. At the same time, we ought to be aware that the source of our conviction in these assumptions does not derive from conditions that objects *überhaupt* (the intentional correlates of our states of experience *überhaupt*) must satisfy, but rather from our cognitive need to have sufficient criteria for forming concepts and beliefs that pertain to the specific character of empirical objects. The crucial difference between the principles of reason and the principles of the understanding

Constitutive epistemic norms both specify how we must think and how nature as such must be; regulative epistemic norms specify what assumptions we must adopt about nature, but the truth of these assumptions cannot ever be strictly verified.⁵⁹⁷

This metaphysical difference in the implications that constitutive and regulative epistemic norms have for *objects* of knowledge corresponds to an epistemic difference that concerns the states of mind of *subjects* who fail to comply with the relevant norms. We can programmatically put this difference by saying that someone who fails to comply with constitutive norms falls short of having objective experience whereas someone who fails to comply with regulative norms may still succeed in having objective experience. Here a question arises: do certain violations of regulative norms *also* mark a violation of constitutive norms? One might hold that someone who posits, within the domain of nature, the transcendental causality of freedom in order to explain a given occurrence violates both a regulative and a constitutive norm: first, she violates the general regulative rule that prescribes that one should strive for the greatest possible extension of empirical knowledge (for in treating a given condition as itself unconditioned, she cuts off the empirical regress from condition to conditioned); second, she violates a constitutive rule, the Second Analogy which states that every natural event must be attributed to a cause whose causality stands under deterministic natural laws. But I think that someone who violates the Second Analogy fails to even get into a position where she would be capable of violating the regulative norm. For in failing to comply with the Second Analogy, one fails in an *epistemically fundamental* way that signals that one does not even partake in the activity of empirical concept and judgment formation; the result being that one fails to get an empirical object into view. In this case one fails to advance to a level of determinate, object-directed empirical thought at which the norm that prescribes striving for the greatest possible extension of empirical knowledge could become even relevant.

Hence, compliance with the constitutive norms of the understanding is a precondition for engaging in the kinds of acts of thought that are governed by, and that can be said to satisfy or violate, the regulative norms of theoretical reason. I now want to sketch some cases where a

is – as Kant never tires of emphasizing in the Dialectic – that the latter provide us with transcendental *knowledge* whereas the former are mere transcendental *presuppositions*. For instance (cf. CPR, A 656-7/B 684-5), reason (via the transcendental principle of specificity) does not determine that natural objects *really are* infinitely diverse (whereas the understanding, via the Second Analogy, determines that every natural change *really is* necessitated by some cause whose causality is governed by some causal law), but only presupposes nature as being so diverse ("indem sie die Natur so reichhaltig voraussetzte") that the thinking subject is obliged to search for instantiations of such diversity. This means that our doxastic state of assent to the proposition that nature is systematically unified ought to be weaker than our assent to the proposition that all natural events are subject to empirical, nomological necessity: the subjective degree of conviction we ought to have in the idea that nature is systematically unified is lower than the degree of conviction we ought to have in the claim that things in nature are governed by nomological necessity. Nonetheless, this degree of conviction is (or, rather, *ought to be*) so strong that it leads us to treat the idea that nature is systematically unified as strictly non-negotiable: we shall never abandon it, whatever our experiences. In the Canon (CPR, A 826-7/B854-5), Kant calls this form of assent 'doctrinal faith' (to be distinguish from practical faith).

⁵⁹⁷ CPR, A 647/B 675: "The hypothetical employment of reason, based upon ideas viewed as problematic concepts, is not, properly speaking, constitutive, that is, it is not of such a character that, judging in all strictness, we can regard it as proving the truth of the universal rule which we have adopted as hypothesis. For how are we to know all the possible consequences which, as actually following from the adopted principle, prove its universality? The hypothetical employment of reason is regulative only; its sole aim is...to bring unity into the body of our detailed knowledge, and thereby to approximate the rule to universality."

thinker violates regulative norms but simultaneously complies with constitutive norms. For instance (1), (cf. CPR, A 654-5/B 682-3) some researchers may emphasize the rule that prescribes that one must look for homogeneity in nature at the complete expanse of the rule that prescribes that one must look for specification in nature, or vice versa. Or, more drastically (2), a researcher may altogether ignore the demand to search for systematic unity and attempt to determine the evidential weight of a perception in an 'atomistic' fashion, under the idea that there is a one-to-one evidential correlation between statements about the world and individual sensory impressions. In a rather different vein (3), an empirical researcher may dogmatically declare that there is no freedom, hence assuming that the conditions of possible experience are conditions of being in general, and thus failing to heed the alert or admonition, which is (as we saw) part of the content of regulative principles, to pay attention to the limits of human understanding. In a still different vein (4), a researcher may strive for systematicity in her cognitive pursuits but believe that such systematicity is a property built into the very nature of reality, thus going entirely against the point of regulative principles which license only 'hypothetical' ('as if') assumptions about the nature of reality. In all these cases, there is a cognitive deficiency that may well have a negative effect on the cognitive pursuits of those subjects who exhibit this deficiency, but those subjects – unlike those who violate, say, the Second Analogy – still qualify as having, at a basic and fundamental level, experience (in general) of empirical objects (in general).

These remarks suggest a complicated and somewhat peculiar, but by no means inconsistent relation between the constitutive norms of the understanding and the regulative norms of theoretical reason (and the power of judgment): while the former are more fundamental than the latter in the sense that violating them leads to destruction of the most basic form of a cognitively significant state that a subject may be in (experience in general), regulative norms nonetheless have a superior kind of authority and dignity in that they, unlike constitutive principles, include in their content a reminder of the sources of human cognition that delimits the boundaries of theoretical knowledge and tempers the pretensions of the understanding. But compliance with *both* constitutive *and* regulative principles is required for securing that we can arrive at the greatest possible extension of empirical knowledge.

Chapter XII: Empirical Belief and the Causality of Nature

In the preceding chapters (IX, X, XI) I discussed various dimensions of Kant's conception of norm-governed epistemic agency, and of the kind of freedom or spontaneity exhibited in such agency. I argued on various occasions (chapter IX, section 6; chapter X, section 7) that Kant has a starkly incompatibilist, anti-naturalistic conception of the freedom we exhibit in our empirical thought: he thinks that we cannot attribute the spontaneity of the thinking, conceptualizing mind to the causality of nature. In this chapter, I want to examine and assess the motivations underlying Kant's view.

1: 'Determination By The Truth'

We can begin by reminding ourselves of the main reasons that may lead one to adopt an incompatibilist, anti-naturalistic conception of *practical* spontaneity. First, the notion that our reasoning toward practical judgments, and hence those judgments themselves, are causally determined by contingent natural causes (such as desires) seems hard to square with the objectivity and necessary validity to which (at least) our moral judgments aspire, or with the notion that these judgments have a kind of rational *authority over* our contingent desires (cf. my argument in chapter V). Second, the notion that our choices are governed by deterministic natural causes seems hard to square with our self-conceptions as imperfectly rational, responsible agents who have the capacity to comply with or to violate practical norms (cf. my argument in chapters VI and VII). *If* our practical judgments and choices are exclusively governed by natural causality, then the processes that bring about these mental states appear to be *non-rational*. This is the distinctively Kantian motivation of the idea that the space of practical reasons cannot be identified with the space of causes.

Now it may seem that these worries dissolve when we focus on our theoretical rather than our practical judgments and attitudes. Concerning the first worry, it does not seem to tell against the validity of our *beliefs about the world* that they are determined by natural causes: it cannot count against the appropriateness of attitudes aimed at representing empirical facts and objects that they are caused by these very facts and objects. Concerning the second worry, our self-conception as believers, as opposed to our self-image as subjects of choice, does not seem to require a capacity to 'act' (i.e., judge) otherwise.

The tendency to presuppose that the worries about a naturalistic, deterministic conception of *practical* agency lose their force when applied to the formation of our *epistemic* states can be witnessed in important contemporary philosophers. Here we can focus on two staunch *incompatibilists* about practical deliberation and choice who adopt a *compatibilist* picture of belief formation. Thomas Nagel writes that "in forming beliefs we may hope for nothing more than to be determined by the truth"⁵⁹⁸, and David Wiggins argues accordingly:

...the libertarian ought to be content to allow the world, if it will only do so, to convey to the free man who has questions that he wants to put and is suitably placed to answer them, how the world is. (2003, p. 96)

If the distinctive aspiration of belief is not to shape the world but to contrive to match or represent it as it is, then I can surely honor the enjoinder in an autonomy of belief untrammelled by the absence of the genuine open alternatives which my freedom requires when I come to deliberate how to *act* upon the world. The world must not dictate how a man should deliberate to *change* the world, but on any view the

⁵⁹⁸ Nagel 1989, p. 116.

world must dictate how the world is. (1969, p. 143)

To be 'determined by the truth' is, I take it, to have one's belief about the empirical world determined by the impact the world has on our sense organs. Since it is the point of belief to truthfully represent the way the world is, the fact that our beliefs are causally determined by the influence the world has on the mind is compatible with our self-conception as thinkers who form objective, world-directed thoughts.

On this picture, belief-formation is an essentially passive matter, responsive to the flow of information that causally runs from states of the world to the subject's mind: as a believer "I simply have to lay myself open to the world in order to let the phenomena put their print upon me".⁵⁹⁹ There is no emphasis here on the relevance of conscious, reflective appreciation of reasons and norms to belief-formation: such appreciation is neither sufficient nor necessary for standing in the causal relations to external objects that determine one to adopt the relevant beliefs. The passive picture *does* support an externalist notion of epistemic normativity: a subject's beliefs are as they ought to be if they are caused – in a reliable manner – by the worldly states that the beliefs are about. This externalist view of belief-formation, and of what it takes for our beliefs to be *well*-formed and justified, summons familiar internalist complaints.⁶⁰⁰ I shall, however, presuppose that the proponent of a causal, externalist conception of empirical knowledge and justification can answer these complaints. In the course of this chapter, I want to examine certain difficulties for the externalist-causal picture of empirical belief-formation that, I think, are largely independent of internalist intuitions.

Before delving into this discussion, I want to notice a few important preliminary points. First, it must be emphasized how congenial the causal, naturalistic account of belief-formation that is implied by the idea of 'determination by the truth' is for a hard incompatibilist conception of freedom and autonomy, such as Nagel's. Nagel thinks that the 'subjective' standpoint of freedom or autonomy is debunked as being based on illusions by the 'objective' standpoint of

⁵⁹⁹ Wiggins 1969, p. 145. Compare G. Strawson's claim that "there is no action at all in reasoning and judging...action, in thinking...is waiting, seeing if anything happens, waiting for content to come to mind, for the 'natural causality of reason' to operate in one" (2004, p. 232). Strawson seems to differ from Wiggins insofar as he thinks that this passive model also applies to practical reasoning.

⁶⁰⁰ One such complaint is that externalist epistemology cannot be considered a normative discipline. Against this, externalists typically argue that their theories do give normative advice: they tell us how one can consciously try to enhance one's chances for *coming to stand* in the appropriate causal relationships, and externalist epistemic norms prescribe how one ought to do this, or specify which belief-forming processes are permissible or mandatory (Papineau 1992; Goldman 1986). A deeper internalist worry is that on the externalist conception, a thinker does not have to be aware of norms or of any reasons for believing; all that matters is that she stands in the relevant causal relations. For an externalist like Papineau, this is no problem: he thinks (1992, p. 2) that "our desire for knowledge derives from our desire to avoid error", and the satisfaction of this desire requires only that one is subject to causal processes that reliably generate true beliefs. But the internalist can respond that we cannot reduce our desire for knowledge to the desire to reliably avoid error: we are reflective creatures to whom it matters that they understand why they are getting things right and that they *can defend* these states *as justified*. The famous counter-examples to reliabilism, such as Lehrer's 1990 'Mr. Truetemp', all exploit the intuition that there is something wrong with a believer who is reliably connected to the truth without any reflective appreciation of his reasons for belief. A typical externalist response to such cases (Goldman 1986, pp. 111-2) involves positing a 'negative' condition which says that in order to be justified in believing that p one must not have reason to believe that one's belief is *not* reliably caused. But even if this takes care of Mr. Truetemp, it still leaves us with the idea that someone who cannot support or defend her beliefs in any way is epistemically justified, as long as she satisfies the relevant causal conditions.

natural science. If it turned out that the objective standpoint is itself infused with remnants of autonomy, then there would be no vantage point from which we can form the kinds of objective explanatory beliefs that expose freedom and autonomy as illusions, and we would get the result that the output of naturalistic explanation clashes with the self-image of the believer who puts forward these explanations. This is why it is important for Nagel to portray the judgments reached from the naturalistic standpoint as themselves susceptible to naturalistic explanations in a way which does not impugn their legitimacy, namely, as caused by the natural world.⁶⁰¹

Second, I want to notice that proponents of the passive 'determination by the truth' model of belief-formation tend to argue that a less passive conception of belief-formation that gives us genuine control over our beliefs would have the unacceptably voluntaristic implication that we can *decide* what to believe.⁶⁰² Since epistemic voluntarism is unacceptable, we must suppose that our beliefs are causally determined by our perception of the evidence. This line of thought presupposes that the attribution of freedom of belief implies that the will is involved in the regulation of doxastic states. It leaves out the possibility of a non-voluntaristic conception of freedom of belief, such as Kant's, on which beliefs are sensitive to the epistemic subject's reflective awareness of what she ought to believe in a way that does not necessarily involve the will. I have argued in chapter IX that this model of epistemic freedom is coherent and non-trivial.

2: The reflective gap between sensory experience and belief

I want to begin my discussion of the deterministic account of belief-formation by reviewing whether it accords with our self-conception as rational thinkers in a very basic kind of case that seems to lend obvious intuitive support for this account. It is, primarily, *perceptual* belief formation that appears to lend itself to the 'determination by the truth' metaphor. One looks through the window, sees that it is raining, and believes that it is raining. Here it seems that one is caused to have a belief state with a certain content (that p) precisely by the state of affairs that this belief state represents as obtaining (p). Furthermore, one might suggest that belief-acquisition on the basis of (veridical) perception is *the* paradigmatic case of acquiring true, justified beliefs about the world. If, in this paradigmatic case there is no difficulty in understanding how being caused to believe by natural states is at the same being afforded with good reason to believe ('causation' is or involves 'rationalization'), one might be able to extend the notion of determination by truth from perception to other kinds of cognition.⁶⁰³

But the notion that we are 'compelled' to adopt perceptual beliefs by the objects of belief does not seem to accord with our self image as thinkers who have beliefs about our environment. We do not think of the formation of our beliefs as events that simply happen to us, and over whose occurrence we have no control or 'saying'. Being affected with sensations by outer objects

⁶⁰¹ It is worth noticing that on this picture it is much to be preferred if our belief formation is subject to a deterministic rather than an indeterministic causality. For if the causation of our beliefs were indeterministic, this would introduce an element of randomness into the picture of how our beliefs evolve that might be incompatible with the notion of 'rational compulsion' or 'determination by the truth'.

⁶⁰² See my chapter IX, section 3 for a closer examination of this tendency and its connection to Hume.

⁶⁰³ I get this line of thought from Brewer: "A person's direct perception of the way things are in the world around him provides a paradigm case of the integration of causation and rationalization constitutive of compulsion in thought by reason, which is essential, quite generally, if the capacities we have been concerned with, for language use, deductive reasoning, and now perception, are to make a genuinely cognitive impact on their subject, in the growth of knowledge and understanding" (1995, p. 17). Unfortunately, Brewer does not spell out how the extension from the perceptual case to the case of, for instance, deductive reasoning is supposed to work.

is, to be sure, an entirely passive matter over which we have no control; but we do not take the beliefs that we may or may not form on the basis of those sensations to be similarly passive. This is (partly) because from our subjective perspective we can always raise questions about whether our perceptual experiences are veridical. We know that it is possible to have non-veridical perceptions of *p* while *p* is not the case (e.g., the stick in the water appears to be bent), and that we are not compelled to believe that *p* in such cases. It is hence possible to acknowledge that the senses present *p* as being the case and yet to withhold assent to the proposition that *p* is the case. This possibility persists in cases of veridical experience, when there *is* a correspondence between the external world and what the senses present as being the case. It is possible to mistakenly take oneself to suffer from a perceptual illusion, or at least to wonder whether one suffers from such illusions. This is because one knows that the presence of an external object is not a necessary condition for having a sensory representation of there being such an object.⁶⁰⁴

It seems possible to generalize from cases in which subjects of perception withhold assent to what the senses present to them to cases where subjects of perception give such assent: the adoption of a perceptual belief quite generally depends on the subject's free assent to what the senses present to her as being the case (see my chapter IX, section 3 for a more detailed defense of this idea). Consider the analogous case of belief based on testimony: being told that *p* by someone else sometimes does and sometimes does not lead one to assent that *p* because quite generally one's assent to *p* depends on one's trusting the source of testimony. Hence, our self-conception as subjects of perceptual belief incorporates the idea of there being what I previously (chapter IX, section 3) called a reflective gap between having a sensory experience and the formation of a perceptual belief. The filling of this gap cannot be attributed to the influence that external objects have on our sense organs.

These points put pressure on the strong assumption that being affected with a sensory experience by external objects forces a doxastic state upon subjects of perception. It may be responded that this strong assumption is not really essential to the picture of perceptual belief-formation implied by the notion of 'determination by the truth'. Maybe the point here is, rather, that with this notion in hand we can make sense of the idea that the natural world puts a constraint on perceptual belief formation that is both causal and rational: if a subject adopts a belief that *p* on the basis of being caused by *p* to have a veridical sensory experience of *p*, the causal impact of *p* on the subject is a necessary *causal* condition of the formation of the belief and a sufficient *normative* condition of the justifiability or rationality of the belief. However, I

⁶⁰⁴ This line of thought may call to mind the 'argument from illusion', which many contemporary philosophers of perception reject. I am happy to grant to the 'disjunctivist' or direct realist, for the sake of the argument, that the objective character of one's perceptual state in the veridical case is wholly different from the character of one's perceptual state in the illusory case. (I am also happy to grant that in some cases where we adopt a perceptual belief on a veridical basis this belief gives us knowledge even if we do nothing to ascertain that we are in the veridical scenario.) But I think that it would be too extreme to demand that we refrain even from saying that it may seem, subjectively, to a person who unknowingly suffers from an illusion that the character of her sensation is as it is in a case of veridical experience. For if we reduce the common 'subjective' factor of experience and illusion to zero, then it might become wholly unintelligible why a person who adopts a belief based on illusory experiences can be as firmly convinced that things are the way the senses present them as being as a person who has a veridical experience. It is worth noticing here that one of the main proponents of direct realism/disjunctivism, John McDowell, concedes that it is possible for a subject to mistakenly take herself to have non-veridical experience and *on this basis* rejects the idea that veridical experience of *p* causally necessitates belief that *p*; see his 1994, p. 26.

think that the notion of a causal constraint imposed by external objects on perceptual belief is too promiscuous to support the idea of a rational constraint. *False* perceptual beliefs are causally constrained by external objects, and this can even involve veridical sensory experiences: a bent stick in the water might cause John to have the experience of a bent stick, and this might set in motion a process that causes John to believe that the stick is straight. *Even true* perceptual beliefs can be unreasonable despite being causally dependent on the occurrence of veridical sensory experiences. Suppose John looks at a round tower from very far away, the tower appears round to him, and John assents to the appearance. Here it seems plausible to say that John ought to suspend belief unless he has gotten closer to the tower, given well-known facts about the extent to which our sensory representation of objects is distorted by distance. These cases suggest that the causal processes that determine our beliefs cannot be regarded as rational or as justified *simply because* they are causally constrained by the impact that the external world has on our senses. The internal processes involved in belief-formation may be set into motion by the impact the world has on our senses, but this does not preempt questions about the justifiability of their product nor, consequently, about the rationality of the processes themselves.

This issue becomes more pressing when we look at beliefs that are based on perceptual evidence but also require some process of reasoning or inference. For instance, one might wonder how the attempt to rationalize beliefs by appeal to the causal relation between belief and fact believed works in cases where one forms a predictive belief about a future fact that does not yet obtain. The traditional response given by proponents of the 'causal theory of knowledge' is that in such cases the fact believed and the belief *have a common cause*.⁶⁰⁵ If a scientist discovers volcanic activity and adopts the belief that the volcano will erupt within a week, the volcanic activity causes both the eruption and our belief about the eruption. But the talk about there being a common cause here is rather misleading, because the sense in which the volcano's activity causes the scientist's belief differs fundamentally from the sense in which it causes the eruption: the volcanic activity is *sufficient for* the eruption in a sense that does not apply to the formation of the scientist's belief. Moreover, citing the volcano's activity as the cause of the scientist's belief does nothing to rationalize her belief: her belief is unreasonable if her perception of the volcanic activity sets in motion a process of ratiocination that is insensitive to an appreciation of *why* the volcanic activity licenses the prediction (e.g., that is based on hasty, scientifically irresponsible speculation). Hence, the fact that the scientist's adoption of the belief is causally constrained by the conditions that produce the predicted event (the eruption) is insufficient to preempt intelligible questions about whether the causal processes that lead from her perception of the evidence to the predictive belief can be regarded as rational.⁶⁰⁶

These considerations show that the deterministic model of belief-formation is not in fact congenial to our 'naïve' self-conception as rational thinkers, and that the idea that the world puts a constraint on our belief-forming processes can at best be taken to implicate a necessary (rather than a sufficient) condition on rational belief. I now want to adduce two further considerations that support the existence of a reflective gap between the reception of sensory data and belief-

⁶⁰⁵ See Goldman's classic 1967, p. 364.

⁶⁰⁶ This point need not disparage the prospects for the 'causal theory of knowledge' as originally stated by Goldman, that is, of an account that says that it is a *necessary* (rather than sufficient) condition for one's belief that p being a case of knowledge that there is "a causal connection between the fact that makes p true [or simply: the fact that p] and" the belief (Goldman 1967, p. 358). Kant himself endorses the condition that empirical knowledge requires an impact of the object of belief on the subject's faculty of sensibility or receptivity.

formation, and that may be taken to put pressure on the naturalistic-deterministic model of empirical belief-formation.

First, notice that a very important sense in which we contribute actively to what we believe about the world emerges when we take into account the fact that our beliefs about the world involve empirical concepts; it is only by using such concepts that we represent things as being a certain way. It would seem that the notion of 'determination by the truth' holds the promise that we can say that the external world determines us, by causing us to have certain sensory experiences, to form and use one set of empirical concepts rather than another. But this promise seems rather empty. In order to arrive at conceptual classification, it is not enough that one is causally affected with sensory experiences by external objects – conceptual classifications aim to tell us what different particular things have in common, and this requires acts of reflective comparison of one's sensory experiences. Such classifications result when various different sensory experiences are related to one another and combined in certain ways. The reflective comparison that leads to conceptual classification of objects and properties is not something that the objects themselves do for us; it must be carried out by the subject.⁶⁰⁷ Moreover, what the senses give us seems to underdetermine our commitment to one among many incompatible conceptual schemes. As Ralph Walker aptly puts it, with respect to some given conceptual classification "there is nothing in the perception that compels us to make this classification rather than some other".⁶⁰⁸ But then it looks as if the determination of our conceptual schemes by external natural causes – i.e., by the objects and states of affairs that impinge on our sense organs – is insufficient to lead us to select a particular set of empirical concepts.

Hence, 'simply laying myself open to the world in order to let the phenomena put their print upon me' does not settle the question of how I ought to conceptualize the objects that affect my senses. If the causal impact external objects have on our sense organs underdetermines what conceptual scheme we employ in our beliefs about external objects, it cannot be said that we are determined to believe one thing rather than another *by the very objects that our beliefs are about*. It must then be something *about us* that leads us to classify objects in one way rather than another. Under the assumption that there is only natural causality, this must mean that it is ultimately the way that our minds, qua parts of nature, are constituted that is responsible for what conceptual classification we end up adopting. But if this has to be conceded, then it seems that we have moved rather far away from the idea that the kinds of natural conditions that cause us to conceptualize the empirical world are the very facts, objects, and properties that our cognitive efforts are directed toward. It seems as if we have moved steadily into the direction of a Humean account of the basis of empirical concept formation, which emphasizes that how we respond to what the senses give us is fixed by the arbitrary 'habits' or 'customs' of the mind.

I want to make this abstract worry somewhat more vivid by looking at one of the most influential 20th century philosophical discussions of empirical concepts, namely, that of Nelson Goodman. Goodman points out that while we do judge, on the basis of sense experience, that emeralds are green, the bare content of those sensory experiences might also be taken to support the judgment that emeralds are grue.⁶⁰⁹ Goodman's Humean sympathies come to light in his

⁶⁰⁷ This is at least part of Kant's point when he says (CPR, B 134) that "combination does not lie in the objects, and cannot be borrowed from them".

⁶⁰⁸ Walker 1978, pp. 128-9.

⁶⁰⁹ See Goodman 1955, pp. 74-75, 103. 'Grue' applies to all things examined before time *t* just in case they are green but to other things just in case they are blue; 'emeruby' applies to emeralds examined for color before *t* and to

claim that the only relevant difference between a classificatory scheme that involves 'grue', 'emeruby', etc. and a classificatory scheme that involves 'green', 'emerald', etc. is precisely that the concepts employed in the latter scheme *are actually* employed or 'entrenched' as a matter of habit. It is unclear whether this appeal to habitual entrenchment is meant to serve as a way of *legitimizing* our classificatory schemes (rather than, in a more authentically Humean spirit, as a way of *explaining why* we use certain concepts rather than others), but to the extent that it is so intended, we seem to have a rather clear instance of a naturalistic fallacy: from the fact that we habitually use a set of concepts it cannot be inferred that we have a right to do so.⁶¹⁰

There are other familiar ways of trying to account for our use of certain classificatory schemes rather than others by appeal to natural facts about us, for instance, by appeal to the idea that concepts such as 'green' and 'emerald', but not 'grue' and 'emeruby', answer to our *interests*.⁶¹¹ It is worth noticing that both the appeal to actual/habitual entrenchment and the appeal to subjective interests imply, in conjunction with the idea that our formation of concepts is wholly subject to natural causes, that it is a *contingent natural fact about us* that we think about the world in one way rather than another. Our thinking about nature in the way we do reflects, ultimately, the way in which we happen to have been naturally constituted so as to habitually project certain concepts or so as to fixate our practical and cognitive interests on certain features. If the laws of nature that are responsible for our psychological or neurological constitution (on which the habits or interests of the mind supervene) had turned out differently, we would – in response to the *same* sensory evidence that, in the actual world, leads us to judge that emeralds are green – believe that emerubies are grue. So, holding fixed the character or constitution of 'external' reality and imagining a different make-up of the features of the mind gives us a different set of empirical concepts. This implies that the way in which we conceptually carve up

rubies not examined before t.

⁶¹⁰ Goodman sometimes seems to suggest that the appeal to actual habitual entrenchment can afford us with a way of legitimizing the conceptual schemes we do in fact use: "...in answer to the question what distinguishes those...features of experience that underlie valid projections and those that do not, I am suggesting that the former are those features for which we have adopted predicates that we have habitually projected" (1955, pp. 96-7). The projection of some predicate counts as invalid "if it conflicts with the projection of a much better entrenched predicate" (ibid., p. 96), where 'entrenched' here means 'rooted in the actual projective practice in which we engage as a matter of habit': "The reason why only the right predicates happen so luckily to have become well entrenched is just that the well entrenched predicates have thereby become the right ones" (ibid., p. 98). This attempt to derive legitimacy from actual employment is extraordinarily puzzling, but maybe one shouldn't take Goodman's talk about validity too seriously – he may be following Hume after all in engaging in a purely descriptive, non-normative approach, albeit of a more refined form: "we are not concerned with describing how the mind works but rather with describing...the distinction it makes between valid and invalid projections. A hypothesis will be said to be actually projected when it is adopted after some of its instances have been examined and determined to be true...the hypothesis need not be true, or lawlike, or even reasonable; for we are speaking here not of what ought to be projected but of what is in fact projected" (ibid., p. 90). I find Walker's appeal to the similarities between Goodman and Kant (1985, p. 24) a bit puzzling. Walker is right to emphasize that Goodman and Kant agree that our conceptual scheme is not fixed by the objects or the contents of perception. But it would be seriously misleading to suggest – and Walker sometimes comes close to suggesting this (see his 1978, pp. 126-30 for an account very reminiscent of Goodman) – that Kant appeals to the actual tendencies of the mind to justify and/or to explain our use of classificatory schemes.

⁶¹¹ For the appeal to the interest-relativity of conceptual classification, see de Sousa 1984 and Hacking 1991. Hacking, in particular, puts emphasis on the *practical* dimension of carving up the world (ibid., p. 113). He does, however, seem to think that this point applies only to *natural kind* concepts, rather than what he considers the cognitively more robust concepts of advanced natural science (ibid., p. 115). See my footnote 612 below.

the world rests on subjective facts about us rather than on objective facts about the world.⁶¹²

The second consideration that I take to put pressure on the 'determination by the truth' model pertains to the relation between belief-formation and evidence. Generally speaking, we can conjecture that there is a close connection between the beliefs a person holds about the world and the evidence that is available to that person; this follows from the popular idea that belief 'aims at truth'. Now, I think that the charm of the notion of 'determination by the truth' rests, to a considerable extent, on a specific, namely, an *atomistic* construal of the evidential relation between sensory experience and empirical belief. Quine famously characterizes the second 'dogma of empiricism as' follows: "to each...statement, there is associated a unique range of possible sensory events such that the occurrence of any of them would add to the likelihood of the truth of the statement"⁶¹³. The model of empirical belief-formation suggested by the idea of 'determination by the truth' holds that the occurrence of a particular sensory experience not only adds to the likelihood of the truth of a particular belief, but both causally and rationally (evidentially) necessitates the occurrence of a particular belief: "A person's direct perception of the way things are in the world around him provides a paradigm case of the integration of causation and rationalization constitutive of compulsion in thought by reason."⁶¹⁴ The 'atomistic' commitment implicit in this suggestion is that there is a causal and evidential connection between (1) a particular worldly fact; (2) a particular sensory experience; and (3) a particular belief that represents this fact as obtaining. The belief is causally and rationally (evidentially)

⁶¹² I do not think that the challenge deriving from Goodman's Humeanism can be answered merely by insisting that certain ways of conceptually carving up the world are intrinsically more objective or 'natural' than others, i.e., that they enable us to 'carve nature at its joints', as Lewis 1983 puts it. Lewis' suggests that the perfectly natural properties will be those that show up in the best ('final') physical theory. But the way natural science arrives at *its* classificatory schemes is, at least to a significant extent, based on perceptual data, and with respect to those data, Goodman's original point persists: the content of the perceptions on which scientists base their classification would by itself support various radically different classificatory schemes. Since scientific classification draws on perceptual data that would (via Goodman) by themselves seem to evidentially support the positing of putatively 'non-natural' properties, the way science carves up the world cannot all by itself provide a criterion for 'naturalness'. One might respond that support for the notion that scientific concepts pick out 'natural' properties derives from the *success* of the concepts used by natural science, but this does not get us very far: not primarily because of the pragmatic upshot of an appeal to success, but because any (say) predictive success with regards to the behavior of things classified as emeralds (or properties such as 'negative charge') also counts as a predictive success with regards to the behavior of things classified as emeralds ('negative charge').

Loewer 2004, p. 186 makes the following suggestion for determining which concepts we should use in formulating our scientific 'axiom systems': "we measure the informativeness of an axiom system so that a premium is put on its informativeness concerning the positions and motions of paradigm physical objects...in evaluating the 'informativeness' of a system we enumerate its proofs by their length and award points for the informativeness of a theorem and extra points if it is about the motions of ordinary objects and then divide by the length of the proof." He then points out that "it seems likely that the gruesome system will have to be a bit more complicated to equal an ungruesome system in informativeness." In other words, the proposal is to rule out Goodman-style predicates by appeal to the fact that they appear to violate the criterion of *simplicity*. Now, a Kantian is bound to find the appeal to simplicity congenial (cf. my chapter XI, section 4). But if we assume that our empirical thinking is solely governed by natural causes, then our reliance on simplicity is itself to be regarded as a 'projective habit'. So, in the naturalistic scenario, the appeal to simplicity can hardly account for the kind of objectivity and non-arbitrariness of our empirical concepts that the Humean story of why we project in the way we do seemed to threaten.

⁶¹³ Quine 1951, p. 38.

⁶¹⁴ Brewer 1995, p. 17.

necessitated by the sensory impact the world has on the subject; and there is no further question to be asked about whether it is rational to assent to this or rather to any other empirical proposition in the face of the sensory experience that one is caused to have.

This picture clashes with the attractive notion that empirical knowledge and justification are *holistic*. The fundamental intuition that is common to the many different versions of holism is nicely summed up by Jane Heal as follows:

The central factor underlying it [holism] is the potential complex interconnectedness of things, both causally and evidentially. The world may present itself to us much of the time as more or less isolated subsystems, further features of which can be inferred on the basis of information about only the current or preceding states of that subsystem. But each such subsystem is, we believe, embedded in a wider spatio-temporal framework which may impinge on it or provide clues about it. (1996, p. 80)

According to holism, the cognitive significance of a particular sensory experience depends on its place within an extensive framework of sensory experiences, empirical concepts, and scientific theories. Surely, Kant's holism differs greatly from contemporary holistic doctrines.⁶¹⁵ But if we accept *some* form of holism, then we cannot suppose that the passive element in perceptual awareness – namely, the reception of sensory input – determines us to assent to a particular empirical proposition. Our assent to some empirical proposition in the view of given appearances rests on our capacities to relate a given appearance to our body of existent empirical knowledge. The cognitive task of properly connecting a given perception to this body cannot be resolved by taking in further sensory input (that would only add to the complexity of the task); hence, in seeking to achieve this task our empirical questions are not settled for us by the way in which we are sensibly affected by external objects. But then there is, once more, a need for an active contribution to empirical belief formation, and this further motivates the idea that there is a 'reflective gap' between the reception of a sensory experience and the formation of belief that cannot be closed by the causal impact that the external world makes on our sense organs. It seems, again, that if our belief formation is solely subject to natural causality, it is a contingent, arbitrary fact about the natural constitution of the mind (or brain) that we devise evidential webs, assign evidential significance to given observations, and confirm or disconfirm particular statements in the light of our totality of sensory experience in the way we do. If our minds had turned out differently, we would use different holistic procedures, and consequently we would end up with different empirical beliefs.

The common point of the considerations adduced in this section is that we cannot easily defuse worries about the notion that our beliefs about the empirical world are produced solely under the influence of natural causality by appeal to the fact that these beliefs are produced in response to the impact that the world has on our sense organs. This impact does not annihilate the need for active contributions of the epistemic subject to belief formation. And if this active contribution is wholly subject to natural causality, this seems to raise a legitimate question about the objective legitimacy of our empirical beliefs. In what way can our *actual* cognitive responses to the impact objects have on the senses be deemed objectively or rationally superior to our vastly divergent cognitive responses in a psychologically alternative possible world where our minds or brains have evolved, and have been 'wired', in a vastly different manner?

It will be noticed that the worries I have expounded in this section concerning the notion that our epistemic agency is solely governed by the causality of nature do not really depend on

⁶¹⁵ See chapter IX, footnote 471 for a discussion of what is distinctive of Kant's holism.

whether this causality is construed in *deterministic* terms or not. That is, these worries do not depend on the notion that we think of ourselves as having a capacity to 'believe otherwise'.⁶¹⁶ Rather, they relate to the *origin* of our empirical concepts and beliefs, and to the character of the processes that control concept-formation and belief-acquisition if (in Kant's terms) 'all causality is mere nature'. This worry is aggravated when we turn from *empirical* concepts to the putative 'pure' or a priori elements in our cognition.

3: Causality: Hume's problem and its contemporary relevance

Kant believes that the attempt to construct a body of empirical knowledge from given sensory experiences requires that there are (pace, say, Quine) *non-pragmatic*, objective epistemic principles that govern the holistic pursuit of knowledge and that can be established independently of appeal to empirical evidence. The thought that empirical cognition stands in need of concepts and principles that are not themselves based on sensory experience is detrimental to the notion of 'determination by the truth/world' that philosophers such as Nagel or Wiggins offer in response to worries about the idea that our empirical beliefs are wholly governed by natural causality. This thought would imply that empirical cognition stands in need of concepts that do not result from the affection of our senses by the external world, and hence cannot be interpreted in terms of Wiggins' idea that as a believer "I simply have to lay myself open to the world in order to let the phenomena put their print upon me".

Now, whatever the precise contours of Kant's picture, it does seem as if some of the concepts that he deems necessary for experience are indeed of fundamental significance to our cognitive orientation and can hardly be attributed to the 'print' that the 'phenomena put upon us'. I shall focus here on the concept of *causality*. This concept plays, after all, a strictly indispensable role in the picture of belief-acquisition suggested by Nagel and Wiggins, or more generally in any reliabilist, causal account of belief acquisition, empirical knowledge or justification. All such pictures must take for granted the legitimacy of the idea that our doxastic states are caused and explained by the existence of those very objects, or by the occurrence of those very events, that these states aim at representing.

Hume famously argued that our beliefs in causal relations are not based on reason because the representation of *causal necessity* cannot be derived from the content of our perceptions. He thus inferred the invalidity of our causal thoughts and concepts from their origin – namely, precisely from the fact that the content of those representations is not derived or copied from the print that the phenomena put on us. In this section, I will argue for the tentative conclusion that Hume's problem is not rendered obsolete within a more contemporary philosophical setting. The conclusion will be tentative because contemporary discussions of metaphysical and epistemological issues surrounding causality are so complex and fragmented that I cannot do justice to the vast multitude of positions within the confines of my study. I can focus here only on a few representative moves that have been made in recent years or decades.

We can begin by noticing that few philosophers have challenged Hume's point that the content of our sense perceptions underdetermines the modally thick content of our conception of causal connectedness. Indeed, it seems that many of those who purport to disagree with Hume about the unperceivability of causality offer a deflationary conception of causation that dispenses

⁶¹⁶ I have discussed the significance of alternative possibilities for Kant's conception of epistemic agency, and for his incompatibilism about epistemic agency, in some detail in chapter IX, sections 4 and 6.

with the element of causal necessity.⁶¹⁷ A notable exception here is Evan Fales, who provides an ambitious defense of the claim that *production*, involving necessity, is an object of direct (non-inferential) perception. Fales focuses on the "type of experience we have when our bodies are pushing and pulling, or being pushed or pulled by, other bodies".⁶¹⁸ This gives us a "feeling of pressure", a "sensation of force".⁶¹⁹ He then claims that what our experience of force directly acquaints us with is not just the the location, magnitude and direction of forces but an asymmetrical relation between cause and effect, e.g., "between self-initiated motion and the passive transmission of motion, as when someone pushes our arm, which in turn pushes a third object."⁶²⁰ But while we certainly draw the distinctions that Evans alludes to, the crucial question is whether the content of our sensations by itself supports these distinctions. The Humean will respond that all we strictly perceive is that our arms move when touched by another arm; she can concede that this is accompanied by a feeling of pressure but deny that having such a feeling implies anything about the *objects* of perception or the relation between them – most importantly, the feeling does not entitle us to the notion that my arm *must* move after being pushed. Fales anticipates this worry :

I imagine the Humean objecting: Why should it be supposed that, for example, a cranially felt pressure will be followed by a cranial motion? What necessity is there in this, since it is admitted that the one does not logically entail the other? But here, I confess, we seem to have reached precisely the point at which little more can be said, beyond asking such a skeptic to re-examine the character of his sensations of force. Certainly, it will be difficult for him to deny that there are similar a priori but non-logical inferences; for example, the inference from something being colored to its being extended; and concerning these inferences also, little further could be said to someone not convinced that they are justified.⁶²¹

Fales here makes a 'partners in crime' move: if the Humean is not willing to accept (1) the necessity of the connection between cause and effect, she is not entitled to accepting (2) the necessity of the connection between being colored and being extended either, for none of these connections – Fales claims – is one of logical necessity. But Hume himself might have said that

⁶¹⁷ This tendency is exemplified by Anscombe 1993 and Ducasse 1993. Both reject a regularity analysis of causation (i.e., both reject the idea that the causal relation presupposes the regular co-occurrence of the cause and the effect). Anscombe pursues an anti-reductivist strategy, i.e., she considers causal efficacy a primitive feature that we directly observe in instances of "cutting,...drinking,...purring" (p. 93). At the same time, she explicitly repudiates "the equation of causality with necessitation" (p. 90). (Stroud 2011, p. 23 overlooks this when he approvingly refers to Anscombe in the context of claiming that we directly perceive causal necessity.) But then the Humean can well respond that if we leave out the idea of causal necessitation, or the idea of a 'thick' modal connection between cause and effect, all that remains is that (say) the butter parts after (or when) the knife touches it. This we can indeed observe, but the Humean never contested that point. If Anscombe insists that we see the *causal efficacy* of the knife's activity, she must do a better job at explaining what causal efficacy minus a certain 'thick' modal connection (minus regularity) amounts to and how the Humean goes wrong in saying that all that perception reveals to us is that one event regularly happens after another. Ducasse differs from Anscombe insofar as he defends a reductivist account, but like Anscombe his understanding of the causal relation is very deflationary: he thinks that the causal relation between two events C and K consists in C's being the only change that occurred in the immediate environment of K immediately before K (p. 127, p. 131). Again, a Humean would not quarrel with the idea that we can observe that much.

⁶¹⁸ Fales 1990, p. 12.

⁶¹⁹ Fales 1990, p. 16.

⁶²⁰ Fales 1990, p. 17.

⁶²¹ Fales 1990, p. 25.

the necessity involved in (2) *is* logical or 'analytic', pertaining to a relation between the ideas of color and body. And more thoroughgoing skeptics about modality, such as Quine, would deny that there is *any* non-conventional necessity involved in (2). Fales argues forcefully against Hume's demand that the necessity in the relation between cause and effect must conform to the standards of logical necessity and non-conceivability.⁶²² But this is only a negative point; it must be supplemented with an account of a potential *source* of a kind of necessity that is non-logical, whose possibility Hume denied; and, moreover, it does not seem as if making room for non-logical necessity would by itself be enough to show that such necessity is directly perceivable.

One might hold that the lack of a worked-out theory of perception on which causal necessity is an object of direct perceptual awareness does not show that we cannot make sense of the direct perceivability of causal necessity. But it should be clear that the Humean attack on the idea that causal necessity is directly perceivable is one of the most influential moves in the history of western philosophy precisely because it seems so intuitively compelling. I would deem it sufficient to respond to any philosophical account that tried to convince me that causal necessity is part of the content of my perceptions that my naïve introspective access to my perceptions affords me no awareness whatsoever of causal necessity (or of other 'thick' modal features). But one can adduce further, less 'dogmatic' considerations here. For instance, it seems that if nomic necessity were directly displayed in perception, then it ought to be possible to tell apart accidental correlations from causal connections on the basis of direct perception. But discovering a genuinely causal rather than a merely coincidental relation between types of events is, typically, not achieved through mere perceptual awareness but may require repeated experiment and theoretical reflection. So someone who wants to hold on to the idea that we directly perceive causal necessity must hold that although causality is a feature of our perceptual content, it somehow lies buried deep within this content and we may need reflection to drag it from there. And this suggestion seems to me to be indistinguishable, for all intents and purposes, from the conviction that do not have direct perceptual awareness of causal relations after all.

Thus, I think that Hume's claim that that causal necessity is unperceivable, and that necessitating causal powers (construed as features of mind-independent reality) are mysterious and unknowable, is still very much worth considering. Now one may respond that all this shows is that we we must be able to make sense of our causal talk *without* invoking the idea of causal necessity. The traditional way of carrying out this proposal is the so-called regularity theory which holds that a causal connection obtains when "*an object [is] followed by another, and where all the objects similar to the first, are followed by objects similar to the second.*"⁶²³ This account – or its contemporary variations⁶²⁴ – makes causation unmysterious because the regular occurrence of one event type occurring after another is, arguably, part of the content of our sensory perceptions (Kant would deny this, but I shall bracket this point here). But the conception of causation as regularity faces a number of grave problems: it has a hard time distinguishing accidental from non-accidental regularities or genuine causes from inefficacious epiphenomenal by-products of a causal process⁶²⁵; and, assuming that causation is very tightly

⁶²² Fales 1990, pp. 18-25.

⁶²³ Hume 1993 (the Enquiry), Paragraph 7 ("Of the Idea of Necessary Connexion").

⁶²⁴ For an example of a 20th century empiricist, non-necessitarian regularity analyses of causation, see Ayer 1963.

⁶²⁵ See Lewis 1973a for a presentation of this difficulty and for a discussion of further worries about the regularity analysis of causation. Baumgartner 2008 claims that most of the standard objections against regularity theories fail, but I do not find his defenses convincing. For instance, I do not see that he has a response to the problem

connected with explanation, it is rather hard to see how the mere fact that all A-type events are correlated with (i.e., followed by) B-type events does anything to explain a particular occurrence of a B-type event.⁶²⁶ So an appeal to mere regularity seems too thin to provide us with a satisfying concept of causation.

A further response to the worry about the unperceivability of causal necessity might be that an interest in this property vanes as soon as we focus on indeterministic (probabilistic) rather than deterministic causation.⁶²⁷ But this won't do, for two reasons. First, accepting the possibility of indeterministic causation hardly dispenses with the actual importance of the idea of physical necessity in our ordinary as well as scientific thought. The indeterministic implications of quantum physics do not seem to tell against the notion that if (e.g.) someone chops off my head I *must* die, or the notion that if I drop my watch in my room it *cannot* hover in the air.⁶²⁸ Second, an appeal to probabilistic causation faces an analogue of the problem Hume raised about the perceivability of causal necessity, for the idea of there being an *objective* probabilistic relation between cause and effect requires the notion of a *real connection* that cannot be analyzed in terms of what we observe; in particular, it cannot be analyzed in terms of regularities.⁶²⁹

As I mentioned previously, the contemporary literature on causation is so vast that I cannot do justice here to all approaches to reduce the causal relation to more 'basic' features. However, I think that none of them (excluding regularity accounts) preempts Hume's problem, for, arguably, *all of them* hold that the representation of causality involves the representation of some objective yet unobservable, *modally thick* connection between cause and effect. This is so for good reason: a 'modally neutralized' account of causation cannot claim any plausibility as an account of our ordinary or scientific causal and explanatory thought.⁶³⁰

that there are many properties that regularly accompany causally efficacious events and which we deem causally irrelevant. He appeals (pp. 5-7) to the idea that (e.g.) singing a song can be eliminated from the conjunction of factors that regularly precede the lighting of a match because it is not part of a 'minimally sufficient' conjunction of factors. But in adjusting regularities to the ideal of minimal sufficiency, we are already departing from a reduction of causal efficacy to mere regularity.

⁶²⁶ See Dretske 1977 and Stroud 2011, pp. 26-9 for elaborations of this difficulty.

⁶²⁷ See Salmon 1984; see footnote 630 below for a brief discussion of his account.

⁶²⁸ For the view that deterministic relations are prevalent at the macroscopic level, see Baumgartner 2008.

⁶²⁹ An appeal to regularities is impossible here, because indeterministic causes are *not* regularly followed by their effects. Moreover, there *is* room for causal necessity in an indeterministic, probability-oriented picture: on such a picture causes are considered necessary and/or sufficient conditions, not for the occurrence of their effects, but for a higher probability of such occurrence; see Schaffer 2001, p. 76. One might try to reduce all such talk to talk about mere *frequencies*, but this would (I think) raise problems analogous to those that plague the regularity analysis. The different suggestion to analyze causal probabilities wholly subjectively in terms of degrees of belief is bound to distort our thoughts about objective causal relations.

⁶³⁰ *Counterfactual theories of causation* in the Lewisian tradition might in part be motivated by the ambition to reduce potentially mysterious causal facts. But what such theories offer as an analysis of 'A caused B' is (cf. Lewis 1973a) something along the following lines: if A had not occurred, B would not have occurred; or if A had not occurred, the chance of B occurring would have been less. And these are modal claims concerning *thick connections* between events whose truth-conditions cannot be drawn solely from sense experience. The Lewisian assessment of counterfactuals requires the idea that some possible worlds are closer/more similar to the actual world than others, and the credentials of such talk seem no less mysterious than the credentials of the talk about causal necessity that originally worried Hume. Thus, Fales (1990, p. 3) aptly diagnoses that abstinence from causal necessity cannot be cured by attention to subjunctive conditionals: "talk of physically possible worlds, and the like, only defers the day of reckoning". Woodward (1990, pp. 558-9) applauds the Humean skepticism and points out that "scientific discussion typically says nothing about metaphysical relations of necessitation". But

Now, there is an influential contemporary *non-reductive realism* about causation which is explicitly anti-Humean in its commitment to causal necessity. It has been put forward, in different forms, by David Armstrong, Fred Dretske, and Michael Tooley.⁶³¹ The core of this approach is the idea that causation presupposes causal laws⁶³² and that a causal law is a relation between universals: if it is a causal law that all Fs are G, this fact consists in the universal F standing in a relation of *nomic necessitation* to the universal G. I share the intuition that causation involves nomic necessitation, and I find the arguments that Armstrong, Dretske and Tooley adduce against deflationist and reductivist views of causality illuminating. However, it is very hard not to feel that their positive account answers the Humean doubt about causal necessity by mere *fiat*. As Theodore Sider asks: "why should some state of affairs in the lofty realm of universals, a fancy relation's holding between two universals, translate into hard facts below, facts about earthly particulars that fall under those universals?"⁶³³ The Humean worry that causal connections are mysterious and unknowable can hardly be defused by an appeal to the existence of unperceivable, abstract entities that are supposed to make (we do not know how) a difference in the concrete spatio-temporal world. One mystery is here being replaced by another. Tooley has tried to address these and related worries, but what he offers remains, to my mind, stipulative.⁶³⁴

his 2003 interventionist theory of causation is (i) centered around counterfactual claims about what would happen in case of counterfactual interventions and (ii) makes liberal use (p. 71) of a notion of what is *physically impossible*. Again, if the Humean program puts a ban on causal necessity, it is hard to see how the modal concepts employed in (i) and (ii) survive this ban.

Another important theory views causation as a 'process-linkage', i.e., as rooted in the existence of a connecting line from cause to effect. It is controversial whether such views are committed to *de re* modality. For instance, Salmon 1984 rejects what he calls the 'modal conception' of causal explanation. However, his point here seems not driven by Humean doubts about modality but by the view that probabilistic causation is more fundamental than deterministic causation; and as we saw earlier, the positing of objective probabilities is, from a strictly empiricist point of view, as suspicious as the positing of objective necessity. When Salmon holds (ibid., pp. 202-3) that causal processes 'carry' probability distributions for different interactions, he should – as a causal realist – mean that these distributions reflect objective connections in nature that do not collapse into mere frequencies. (For the view that the processes that process linkage theories focus on are best characterized as lawful sequences, see Schaffer 2001, pp. 77-8.) Moreover, Salmon's attempt to distinguish causal from pseudo-causal processes involves (pp. 147-8) an appeal to what 'characteristics' a process *would* manifest over time in the absence of certain conditions; and his talk about 'marks' that are 'transmitted' by causal processes is hardly based purely on observables. There are other versions of the process theory, some of them more austere than Salmon's; I can here only register my skepticism about the attempt to devise such theories without appeal to thick modalities.

⁶³¹ See Armstrong 1983; Dretske 1977; and Tooley 1988.

⁶³² For Armstrong and Tooley, this is compatible with a qualified 'singularism', i.e., a rejection of the Humean supervenience thesis that causal relations are logically determined by laws plus non-causal properties. This issue is not crucial for our purposes.

⁶³³ Sider 1992, p. 261.

⁶³⁴ Thus, Sider 1992, p. 262 calls Tooley's method for identifying the necessitation relation "solution by stipulation": Tooley simply postulates a number of properties that the necessitation relation must have *by definition*, and then invokes the Ramsey-Lewis account of defining theoretical terms (1988, pp. 78-80). Elsewhere (1993, pp. 4-5), Tooley suggests that Hume's "crucial handicap" was that he did not have a proper method for providing a meaning for theoretical terms referring to unobservable entities. But I find it unsatisfactory to imbue the Ramsey-Lewis method with the power to answer Humean skepticism all by itself. The Humean worry pertains to what goes on *before* the Ramsey-Lewis method is invoked: namely, the selection of properties that the nomic necessitation relation *is intended to have by the subject* and that we can neither know nor understand through observation. That an appeal to the Lewis-Ramsey method is neither here nor there as far as the issue of causal necessitation is concerned is proven by the fact that Lewis uses it (1973b, p. 73) to account for a theory of law-

It thus seems to me that Hume's problem about causality persists in a contemporary setting, although (to repeat) I do not take myself to have shown conclusively that the problem cannot be resolved, or preempted, with the resources of one or another contemporary theory. Now, if Hume's problem persists, so does the famous question that he raises in the context of posing this problem: what *determines us* to believe in causal relations, reason or habit? Hume's equally famous answer to this question is that *since* causation is not some property that we can detect through the senses as a result of our interaction with the 'external' world, our idea of causality is forced on us 'from the inside' – as a result of the internal mechanisms of the mind, the principles of 'custom' or 'habit'. We cannot help but believing that events are causally connected because it is in our psychological nature to believe this; what explains our causal beliefs are non-rational, psychological (or neurological) laws and processes. One can think of a similar Humean account of our commitment to other 'pure' concepts that seem central to our cognitive orientation and that are not derived from observation (i.e., copied from the content of our perceptions of external reality), such as the ideas of systematic unity or simplicity.⁶³⁵

It is worth noticing that a causal theory of knowledge does not help here, for it presupposes precisely what the Humean challenges: i.e., it would have to claim that causal powers and connections are features of reality to which we stand in a reliable, causal connection; and this dogmatically eliminates, by *fiat*, Hume's point that the evidence that the external world provides us with via affecting our senses significantly underdetermines the notion that there are causal powers and reliable causal connections. The external world does not dictate to us the belief that certain conditions are sufficient (or raise the probability) for the occurrence of some subsequent event. If we are determined to believe in causal necessity (or its probabilistic counterpart), and if all causality is natural causality, the dictate must come from our own psychology; and if our causal beliefs are the upshot of our contingent psychological (or neurological) habituation, the processes that lead us to adopt those beliefs seem to be non-rational. As Hume puts it, "when the mind...passes from the idea or impression of one object to the idea or impression of another, it is not determined by reason, but by certain principles, which associate together the idea of these objects, and unite them in the imagination".⁶³⁶

Contemporary empiricists or naturalists tend to be more permissive than Hume in their conception of the objects that our perceptual states relate to. They reject the Humean claim that the immediate objects of our perceptual awareness are fleeting sense-data. But at least with respect to concepts such as causality, it is not clear if this relaxation of the Humean strictures is sufficient to alleviate the worries raised by Hume's theory of concept acquisition, as long as we

likeness devoid of any commitment to natural necessity.

⁶³⁵ Simplicity and unity are of special relevance here because these ideas are invoked by those who try to distinguish between laws and accidental regularities without appeal to nomic necessitation. They hold that "a regularity is a law iff it is a theorem of the best system", and the best system is "one that strikes as good a balance as truth will allow between simplicity and strength" (Lewis 1994, p. 478); or they hold that "to say that a regularity is a [law] is to say that it can be derived from the best system of the world. But this entails that it can be unified by connecting it to the other regularities implied by the best system" (Loewer 2004, p. 189). Now both Lewis and Loewer concede that on this account the property of being a law is "partly constituted by psychological factors" (Loewer, *ibid.*, p. 191; cf. Lewis, *ibid.*, p. 479); but they seem to have in mind here mainly psychological variations in interpretation of simplicity or unity. The more radical Humean psychologism I have in mind pertains to the very ideas of simplicity and unity. It points out that our cognitive commitment to these ideas is ultimately driven by arbitrary and contingent, non-rational psychological causes and conditions.

⁶³⁶ Hume 2009 (the Treatise), Book 1, Part 3, Section 6.

hold onto the essence of Hume's conception of the human mind and consider our thought about nature to be determined by contingent psychological (or neurological) habits.

4: Causal thought and the causality of nature

In the preceding section, I argued that Humean worries about the legitimacy of our use of the concept of causality persist in a contemporary setting. Barry Loewer aptly puts the kind of worry I am concerned with (or a close analogue thereof) as follows:

On the one hand, causality seems to be so intertwined with so many of our concepts (indeed, with the concept *concept*) that if it fails to refer then most of our thoughts would also fail to refer. On the other hand, look as hard as one might, we just don't find causal relations among the fundamental properties of physics or in the dynamical laws of physics. So we would be in a dilemma of either rejecting our conceptual scheme or rejecting physicalism, at least in its austere formulation.⁶³⁷

The concept of causality – as involving some thick modal content that goes beyond the idea of a regular co-occurrence of events – seems so fundamental to our cognitive orientation that an eliminativist stance with regard to this concept is not a real option⁶³⁸; but the best scientific theories provide no evidence for the existence of causal relations. What Loewer should have added here that is that the very idea of 'fundamental properties' or 'dynamical laws' of physics seems to presuppose the legitimacy of our nomic and modal notions.

From a Kantian perspective, the difficulty exposed by Hume stems from the ambition to legitimize modally 'thick' claims (about the way things *must* be) in respect to a kind of reality that is wholly independent of our cognitive faculties. If we cannot legitimize the notion of causality on the basis of (ordinary or scientific) experience, then we might legitimize it by showing that employing this notion is possibility of objective experience. Kant argues that if we accept that the character of the objects of our experience is dependent on (shaped by) our cognitive faculties, then we can defend the principle that the concept of causality is applicable to objects (considered as objects of human experience, and hence subject to the conditions of human experience). More generally, on the Kantian approach concepts such as causality (or simplicity or unity) are not taken from nature but imposed on nature by us.⁶³⁹ Even if one may not agree with the precise details of Kant's theory, its contours seem to provide an attractive strategy for defending the legitimacy of a concept that is as indispensable as it is mysterious.

Here I agree with Patricia Kitcher, who argues that if one takes seriously Hume's challenge but also rejects a regularity theory of causality, one may be led to a Kantian conception of causality as mind-dependent, especially since (as we saw) the concept of causality is absolutely fundamental to a naturalistic or externalist epistemology:

How can epistemologists appeal to causal relations to explain knowledge if they have no theory of our

⁶³⁷ Loewer 2004, pp. 199-200.

⁶³⁸ For such eliminativism, see Russell 1992.

⁶³⁹ The resulting *partial* mind-dependence of our causal claims and laws cannot *per se* be considered a reason for rejecting the Kantian program, for, as we saw, contemporary naturalists such as Lewis and Loewer are committed to such mind-dependence as well. See footnote 635 above. Loewer, following Lewis, points out (2004, p. 191) that in his framework the mind-dependence of causal laws (which results from the subject-dependence of criteria such as simplicity) is compatible with a robust realism; in particular, "which generalizations are the laws is mind-independent". But Kant can say something analogous. As we saw (see chapter X, section 7 and chapter XI, section 2), Kant holds that the spontaneous mind imposes only the general abstract feature of causality on nature. It does not determine what the specific laws of nature are.

epistemological access to causal relations? Contemporary epistemologists may try to finesse worries about access to causal relations by assuming that it will be possible to work out some refinement of Hume's 'positive thesis'...the 'regularity' account. Unlike powers and grounds, regularities seem to be cognitively accessible. But... the problem is that it is not clear that Hume's positive thesis about a pattern of regular succession, C's followed by E's, makes any sense except in relation to some creature or device for whom or for which that is a cognizable pattern. In which case, dodging Hume's criticisms about mysterious or unanalyzed notions of causation may lead to an analysis where 'causation' is (partially) 'mind-dependent.' Hence, Hume's celebrated critique seems to confront causal epistemologists with an awkward choice: carry on with an unanalyzed (and so suspect) notion of cause, or acknowledge that causation may turn out to be (partially) 'mind-dependent'....⁶⁴⁰

Let us assume that causation is indeed, in one way or another, mind-dependent. What I want to investigate is whether this assumption can contribute to a legitimization of the concept of causality under the further assumption that our mind is part and parcel of the natural order, i.e., wholly and solely governed by *natural* causes. As I argued previously (cf. chapter IX, section 6; chapter X, section 7), Kant denies this further assumption. I think that it would not be an exaggeration to suppose that his conception of 'pure' epistemic spontaneity was crucially shaped by his worry about the Humean debunking of the idea of causal necessity as a 'fiction' that we cannot help but adopting under the influence of non-rational psychological processes.⁶⁴¹

Now, one may ask: what would be so bad about the notion that we come up with and employ the concept of causality solely as a result of the workings of our 'inner' natural constitution? One might hold that it is one thing to say that the concept of causality comes from our own psychology (or neurology) and quite another thing to say that the concept of causality lacks objective validity; maybe Hume was right about the first but wrong about the second point. We might, against Hume, be able to offer a rational justification for thinking in causal terms, while agreeing with Hume (and the contemporary naturalist) that thinking in causal terms is a function of our psychological or neurological constitution. Some interpreters of Kant, such as Patricia Kitcher, even claim that this is the position that Kant himself adopts, or at least should have adopted if he had freed himself from his anti-psychologist prejudices.⁶⁴²

I think that Kant himself would protest, first and foremost, that this picture, in which the concept of causality derives from our contingent empirical psychology (or neurology), is incompatible with the a priority of the concept of causality and (hence) with the necessity of the causal principle (C), 'every alteration has a cause', which employs this concept. This might seem surprising: isn't Kant, among all philosophers, committed to the idea that the origin of a concept is irrelevant to its validity? I have discussed this point in some detail on a previous occasion (chapter V, section 1; pp. 92-3). In a nutshell, my response is that Kant thinks that we cannot justify a priori concepts by appeal to their empirical origin because he holds that if the empirical origin of a concept that purports to have necessary, universal validity could be demonstrated, the

⁶⁴⁰ Kitcher 1999, p. 418.

⁶⁴¹ Walker claims (1978, p. 6) that "there is no evidence that [Kant] took Hume's doubts seriously or even gave them much thought". But there is substantive textual evidence, even apart from speculations about the role that Hume played in the development of Kant's thought, for supposing that Hume's skeptical challenge to our reliance on the notion of causality, and his attempt to explain causal beliefs by appeal to psychological custom, significantly shaped Kant's insistence on the freedom of our thought from psychological causes: Hume's subjectivist account of causal belief is mentioned explicitly three times in the First Critique (CPR, B 5, B 128, A 765-6/B 793-4) and figures prominently in the Second Critique as well (KpV, 5: 12-3; 51-3).

⁶⁴² See my chapter X for a discussion and rejection of this interpretation.

concept would thereby be invalidated.⁶⁴³

One might object that there are obvious counterexamples to this position: the apriority and necessity of the claim (B) that all bachelors are unmarried male adults does not at all seem to be affected by the idea that it is a contingent fact about us that we form and use this concept. But examples like this are unhelpful for the purposes of our discussion. For the difference between (B) and the causal principle (C) that (all) events are causally related to preceding events is that the former is (in Kant's terminology) analytic whereas the latter is synthetic. With regard to analytic principles, their apriority and necessity follows from the internal structure of the relevant concepts. Such facts about conceptual structure, if they exist, will indeed be unaffected by the contingent origin of the concept (even Hume acknowledged 'relations of ideas'): how we got the concept does not matter, as long as there is a discernible fact of the matter about what marks one must represent if one represents the concept at all. But notice here that *precisely because* the origin of concepts that support analytic necessities does not matter, a concept that supports such necessities does not, per se, have any cognitive significance: even a concept that has its origin in our imagination and that does not apply to any real objects at all can be a source of analytic necessities (consider: 'All unicorns have horns'; see Kant's remarks about 'invented concepts' at CPR, A 729/B 757). While the concept of 'bachelor' does apply to objects, it is unclear what cognitive significance the idea that all bachelors are unmarried male adults is supposed to have, precisely because this claim is rooted in an arbitrary linguistic convention. Kant's point about the link between non-contingent/non-empirical origin and a priori, necessary validity relates only and exclusively to cognitively significant, synthetic claims about objects. The idea (C) that (all) natural events have necessitating causal grounds purports to be a cognitively significant claim about real things, which cannot be verified either on the basis of perception or by appeal to the

⁶⁴³ See EEUK, 20: 239 (a passage I quoted twice before); compare KU, 5: 182 and R 5636, 18: 267-8. In this context we must also briefly consider the logic of Kant's response to Hume. Kant says, in a preliminary response to Hume's attempt to give the concept of causality a foundation in empirical laws: "For whence could experience derive its certainty, if all the rules, according to which it proceeds, were always themselves empirical, and therefore contingent? Such rules could hardly be regarded as first principles" (CPR, B 5). Now, one may object that for Kant a concept is a priori if it is not derived from experience, and a concept is not derived from experience just in case it is not abstracted from the content of perceptions (cf. Warren 1998, p. 215); if a concept is produced by internal empirical (psychological or neurological processes) that do *not* copy the content of perception (as on Kitcher's account), these conditions are satisfied. But Kant's notion of a concept's being 'derived from experience' is broader than 'abstracted from the content of perception': he says that Hume tried to derive the concept of cause from experience (CPR, A 94/B 128), and Hume obviously did not think that the content of this concept is abstracted from the content of our (primary) sense impressions. Hume's account counts, for Kant, as a 'derivation from experience' because it attempts to find the origin of the concept of cause in something that we can empirically observe, namely, the habits of the mind that are governed by empirical laws. Henrich 1976, p. 105 claims that one can assert the a priority of a rule while bracketing questions about its origin, as long as (i) the rule can be 'thought a priori' and (ii) 'allows for transition from any state of consciousness to any other'. But the two conditions he posits here are quite puzzling. Regarding (i), it seems to me that Kant would insist that we qualify as thinking a rule a priori just in case the concept that we employ in our thought has a non-empirical origin, i.e., neither derives from the content of empirical intuition nor from the workings of the empirical imagination. Regarding (ii), Kant would worry precisely that if our synthesis of representations were subject to rules that are imposed on us by the contingent constitution of the empirical mind, then the rules could in principle (depending on how we happen to be constituted) cause the transition from *any* state to *any* other – that is, how we happen to synthesize representations would be completely arbitrary. It should be noted that in a later essay (1989, pp. 35-6), Henrich explicitly rejects the independence of questions of validity from questions of origins (see footnote 214 above).

internal structure of concepts – and here the origin of this claim cannot simply be dismissed as irrelevant, for the idea that we come up with the concept of causality and affirm (C) only because our internal psychology happens to work a certain way seems hard to square with the idea that (C) states a necessary truth about the things represented by the concept, things that are independent of our contingent psychology. Moreover, there is a further reason why the appeal to analytic claims is unhelpful: it is notoriously unclear how it can be shown that concepts indeed have the internal, analytic structure that can support genuine necessities. Even the best candidate claims, such as 'An unmarried male adult is a bachelor', can – it seems – be doubted (the pope is an unmarried male adult; but is he also a bachelor?). Moreover, if the candidate concepts have cognitive significance, i.e., apply to real objects independently of our inventions or conventions, they must be derived from experience, and in this case one might argue that the strict universality required for genuine necessity can never be confirmed on the basis of conceptual analysis (a point that Kant comes close to endorsing at CPR, A 728/B 756). Kant thinks that a few analytic points can be established with regard to pure concepts such as 'substance' and 'cause', but the cognitive significance of those points depends on whether the concept has objective validity, and this is not something we can take for granted independently of considerations about origin.

Now, one might propose that in order to account for the validity of non-analytic necessity claims, we need only find a 'synthetic analogue' that can play the role which internal conceptual structure (allegedly) plays in the case of analytic necessity claims, namely, some fact of the matter that justifies the relevant necessities regardless of origin. If the question is whether the processes that lead us to come up with and to think in terms of the concept of causality are rational, it is – one might hold – irrelevant whether these processes originate in our empirical constitution. All that matters is whether we can, against Hume, come up with an objective justification of the concept of causality: for instance, whether we can show that applying this concept is a necessary condition of objective experience.

In response, I want to suggest that Kant's worry takes the following form. First, note that whatever reasons there are for thinking that events are causally connected, they must be independent of how we actually think as a matter of empirical fact. Even if our minds were so constituted that we often, or maybe even always, fail to represent events as governed by empirical laws and necessity, it would be the case that there are compelling reasons for representing events as so governed. Note, secondly, that if the kind of mental activity that leads us to form and apply the concept of causality is governed by natural causality, our thinking in causal terms depends on the actual empirical constitution of our minds (or brains). These two points jointly imply that if our thinking in causal terms is solely governed by natural causality, it is a lucky coincidence that the empirical constitution of our minds has turned out to be such that we happen to think in the manner we have reason to think, i.e., in causal terms. Moreover, *that* we happen to think in this manner is unrelated to the fact that we have reason to think in this manner, because the evolution of the empirical constitution of our mind which is responsible for our thinking in causal terms occurred independently of facts about what there is reason to think. So if our thinking in causal terms is solely governed by natural processes that depend on the empirical constitution of our mind or brain, we *would* think in causal terms *even if there was no reason* for thinking in causal terms. Hence the fortunate, accidental fact that these processes happen to lead us to think in ways that accord with what there is reason to think does not render them rational processes, and does not render the influence they exert on our thinking a rational

influence. So if our thinking in causal terms is governed by natural laws just like the movement of the planets is governed by the law of gravity, it follows that whether or not there is objective reason to think in those terms is completely irrelevant to whether or not we think in those terms; the 'blind' processes that are responsible for our use of the concept of causality are insensitive to whatever the potential rational credentials may be that speak in favor of employing this concept.

One might try to block this line of thought by suggesting that the natural processes that led to the evolution of the psychological/neurological features that determine us to posit 'thick' causal connections were not in fact independent of there being objective reasons to posit such connections, but were (or are) themselves responsive causally to underlying, independent reasons for thinking in causal terms. The question, then, is how this causal responsiveness of empirical mechanisms to reasons for thinking in causal terms is to be understood. One – and, it seems to me, the most natural – proposal is this: in the course of the evolution of our patterns of thought (or of the neurological mechanisms underlying these patterns), evolution selected for our mind (or brain) to become wired in such a way that it tracks real, objective features of our environment, and our positing of thick causal connections is based on our sensitivity to such features. But this proposal raises the question what this 'tracking' is supposed to consist in, or how it is supposed to occur. And this is Hume's question in new disguise. Under the reasonable assumption that perception is the only device we have for tracking real features of the natural world, and under the assumption (which I examined earlier) that the content of our perceptions underdetermines the content of our concept of causality, it follows that we cannot view the natural evolution of our mind (or brain) as selecting for a tracking of causal features.

One might appeal here to inference rather than perception: don't we track causal connections by forming firm expectations about the future on the basis of past experience, expectations that turn out to have been warranted after all? Here two points are relevant. First, the fact that our expectations about the future have been warranted does not at all support the idea that we have been tracking causal connections: strictly speaking, all that can be said to have been warranted, thus far, is our conviction that certain correlations between types of events that we have observed in the past continue to hold, and this can be conceded in conjunction with the complete denial of the idea that there are any 'thick' modal connections between events (consider 'regularity theories', which are entirely based on the observation of repeated correlations). Second, under the assumption that our patterns of thought are the upshot of contingent empirical causes and conditions, our inferential dispositions – and in particular the disposition to infer the existence of causal laws and necessities on the basis of experience of repeated correlations between types of events – are themselves contingent, in the sense that if the natural history of the human brain had been different, we would not have become disposed to engage in the inferential practices that we actually engage in, even though our perceptual evidence would have been exactly the same. Hence, that we draw those types of inferences can be explained (as it is in Hume) solely by appeal to contingent psychological habits, and without any mentioning of real causal necessity or law-likeness – precisely because, once again, the conclusions we draw about external reality through causal and inductive inferences are causally and evidentially underdetermined by the 'input' we receive from external reality. Thus, the appeal to inference *rather than perception* cannot substantiate the idea that we are tracking necessary causal connections and laws, if our actual inferential practices are the upshot of contingent internal

features that are not derived from, and do not co-vary with, the character of external reality.⁶⁴⁴

Now the idea that evolution selected for the mind to become wired so as to posit thick causal connections can also be understood in a very different way, namely, in terms of the survival value of thinking in causal terms.⁶⁴⁵ How exactly this would work is a bit obscure, but let us grant that our species' probabilities for survival (or well-being) are enhanced by thinking that events are causally connected. The problem is that this appeal to survival-enhancement does not confer the *right* kind of justification on our thinking in causal terms. The justification we are interested in pertains to the question of whether we are *epistemically* justified in thinking that there are 'thick' causal connections between events; the idea that our practice of positing such connections is conducive to our survival (as a species) is beside the point at issue.⁶⁴⁶

My above line of thought was to the effect that if our mind is exclusively subject to natural causality, it follows that even if there were some objective justification for thinking in causal terms, the thought processes by which we posit causal connections could not be considered rational because these processes do (or would) occur completely regardless of whether or not positing causal connections is objectively (epistemically) justified (since the neurological mechanisms that are, on the naturalistic picture, responsible for the occurrence of these processes are, or were, not designed in response to any objective epistemic reasons). It can be noticed here that the idea that we would think in causal terms *regardless of* whether doing so is justified is close enough to Hume's claim that our thinking in causal terms is not based on reason. I take Kant to agree with Hume about the idea that if it had to be conceded that the concept of causality originates in contingent features of our empirical psychology, it would follow that the processes that lead us to form the idea of causal necessity are non-rational. Kant insists on his strong notion of the a priority of this concept partly because he thinks that it is only if our causal thought is not determined by contingent empirical (psychological) causes that we can think of ourselves as rationally or objectively justified in forming and using this concept.⁶⁴⁷

⁶⁴⁴ Armstrong 1983, p. 73 suggests that "...the postulation of laws is a case of inference to the best explanation. The thing to be explained is the observed regularity of the world." Now, it is far from clear whether positing causal laws, or a relation between unobservable universals (which, as we saw, the Armstrong-Dretske-Tooley account requires) really is a good or the best way to explain the observed regularity of the world. But whatever we may think about this, the fundamental problem I have (following Hume) been belaboring remains: positing necessary laws is surely *evidentially* underdetermined by the observation of regularities; nothing in the content of our perceptions imposes an idea of necessitation (not to mention all the stuff about universals) on us. Hence, these ideas are – given a wholly naturalistic view of the mind – a product of our contingent psychology. A mind that evolved from a different natural history and that was subject to different natural laws would not be prompted to respond to the observation of regularity by positing necessitation relations. That our mind is so prompted can thus hardly be considered as the basis for a rational inference.

⁶⁴⁵ The idea that our thinking in causal terms, or our affirming the synthetic a priori judgment that every event has a sufficient reason, can be justified by appeal to survival-value rather than epistemically is endorsed by Nietzsche 1966 with characteristic rhetorical force: "the time has finally come to replace the Kantian question 'How are synthetic judgments *a priori* possible?' with another question, 'Why is the belief in such judgments *necessary*?'— that is, to understand that for the purposes of preserving beings of our type we must *believe* that such judgments are true, although, of course, they could still be *false* judgments!" (Beyond Good And Evil, Par. 11)

⁶⁴⁶ In this respect, the appeal to the pragmatic (survival-enhancing) value of thinking in causal terms is on a par with Strawson's 2003 attempt to defend our practice of holding people responsible by appeal to practical reasons.

⁶⁴⁷ See CPR, B 167-8, and my chapter X, section 5 for a discussion of Kant's point that if we had to concede that our thinking in causal terms were the upshot of an "arbitrary subjective necessity" so that we could not help but think in those terms as a matter of psychological habit, we would be thrown into skepticism. Notice here that Kant's position is even more radical than the one I am expounding: he holds that we could not take ourselves to be

Kant's intuition here can be supported by the following thought experiment. Suppose that we are determined to posit 'thick' causal connections by our neurological 'hardware'. Suppose, further, that tomorrow a drastic cosmic change occurs that alters the neurological laws that – ex hypothesi – led us to posit thick causal connections up until now; now our neurology is such that we no longer posit thick causal connections. Hence, our causal judgments *in the actual world* are determined by processes that are completely insensitive to whether or not these judgments are supported by adequate reasons: these processes are a mere function of however our mind (or brain) happens to be wired as a matter of contingent empirical fact. It is, I think, this point that Kant has in mind when he says, in his review of the fatalist Schulz:

...even though he [Schulz] did not concede this to himself, he presupposed that the understanding has the capacity to determine its judgment according to objective reasons that are valid at any time, and does not stand under the mechanism of merely subjectively determining causes, which may change subsequently; he assumed freedom of thought, without which there would be no reason.... (RezSchulz, 8: 14)

In response to these worries, one might concede that the processes that lead us to think in causal terms are non-rational, but insist that this does not matter: what matters here, one might say, is whether the concept of causality *can* be given an objective justification from a third-person theoretical perspective, that is, from a viewpoint that is external to the 'blind' use of the concept.⁶⁴⁸ But here the question is why this matters at all, and to whom it is supposed to matter. If the actual use of the concept of causality is completely divorced from thoughts about the legitimacy or validity of this concept, then it is hard to see what the putative objective justification for thinking in causal terms is supposed to be doing for those who apply it. If a verdict delivered from a perspective external to our cognitive practice confers legitimacy on this practice, and if has to be conceded that the practice would continue in exactly the same manner even if the verdict turned out to be less conciliatory, it seems that the verdict that our practice is objectively justified has no bearing on (the rationality of) this practice. But then it is not clear why we or why anyone should care about whether or not we are justified in using the concept of causality. The issue seems utterly idle. It seems that we who use the concept should either be quietists about its legitimacy; or, if we find such quietism unacceptable, we should continue to worry about the legitimacy *of the use we make* of the concept quite regardless of there being a putative justification of the concept, for this justification is wholly external, even alien to our use of the concept. In the envisaged scenario, the sense in which we are 'justified' in applying the concept of causality is similar to the sense in which a computer (in Kant's terms, a 'turnspit') programmed to track the temperature is 'justified' in processing data in a way that indeed tracks the temperature. In both cases, the way in which the mechanism operates is contingent and dependent on the manner in which its internal structure was formed or set up, and it is independent of whether or not there are any reasons for processing the data in the relevant manner. The verdict that processing in the relevant manner is 'appropriate' is merely external and does not have implications for the rationality of the process or its 'agent'.⁶⁴⁹

The computer-analogy raises a further question: namely, what legitimizes the 'external

rationality justified in our causal thoughts even if our minds were wired to think in causal terms by a rational, divine intelligence who benevolently wires our minds in such a way that we infallibly get things right, rather than (as I am supposing) by rationally arbitrary evolutionary processes.

⁶⁴⁸ Both Longuenesse and Kitcher attribute this strategy to Kant; see chapter X for discussion and criticism.

⁶⁴⁹ For further discussion of the computer case, in a more interpretive context, see chapter X, section 6.

verdict' concerning the appropriateness of our causal thought? The verdict that confers *appropriateness* on a computer's processing of data draws on two facts: (a) the computer has been intentionally set up by an external agent to serve a certain purpose (say, to calculate the temperature); and (b) the computer's actual operations do serve this purpose. Now, (a) (and thereby (b)) has no analogue as far as our causal thought is concerned; if the fact that we think in causal terms depends on the contingent empirical constitution of the mind or brain, the evolution of this constitution was not purposive, or guided by intentions, in anything but a metaphorical sense (unless Creationism is true). So we cannot say here that our causal thought is appropriate or justified by reference to some intention or purpose that is itself the cause of our causal thought. We might try to argue that thinking in causal term is a necessary condition for object-directed thought; in this manner, we might try to show that our minds happen to have evolved in such a way that they operate in accordance with necessary conditions for cognition of the empirical world. But notice that in trying to show this, we are in exactly the same position as those whose practice of thinking in causal terms we are supposed to be justifying. The idea of 'objective thought' or 'cognition' that is supposed to be doing the justificatory work is as little an observable or inferable feature conveyed to us by our interactions with the world as the concept of causality itself, and so our commitment to both our concept of cognition/objective thought and to our concept of causality comes *from within*, namely, from the internal set-up of the mind, rather than from the outside; the external world does not determine us to think about cognition in the way we do (or to use the concept of cognition at all). The mental processes that lead us to form and use our idea of cognition are of the same type as the processes that lead us to form and use the concept of causality; and hence if the latter are non-rational, so are the former. That we happen to think about objective thought or cognition in the way we do, or at all, is just another 'custom' that is, ultimately, a function of the way in which our minds happen to have evolved. *There is no way to break out of the box here*, that is, to leave behind the way of thinking that is imposed on us by the contingent set-up of the mind that controls our habits of thought, under the assumption that the mind is just another contingently structured piece of nature. (By contrast, the perspective of the engineer on the computer is a *truly external* outlook.) Hence, the very idea of an 'external' perspective, *conciliatory or not*, on the cognitive practice involving the concept of causality is rather elusive, because any such perspective is already 'infiltrated' by the very modes of thought we are supposed to be justifying or, in Hume's case, debunking.⁶⁵⁰

⁶⁵⁰ The worry about Hume's attempt to adopt a 'non-conciliatory' external viewpoint toward our cognitive practices is this: he (1) takes the human mind to be an object of naturalistic study and seeks to explain the origin and genesis of its fundamental cognitive ideas, such as causality, which Hume thinks he has shown to be fictitious; but (2) in conducting this study Hume cannot get out of the 'human box', namely, he cannot escape those very habits of thought he is supposed to be investigating. In investigating and explaining those habits he must himself make use of the concept of causality, i.e., he must appeal to conditions that produce (mental) events. I find the most illuminating formulation of this worry in Stroud 2006, pp. 348-9 (my emphasis): "What is true of the naturalistic study of both animals and human beings is that the study is carried out in each case by human beings. That presents no difficulty for our understanding whatever we find to be true of animals. *But when human beings are both the agents and the objects of naturalistic study, what is found to be true of the objects studied must somehow also be understood to be true of those of us who conduct the study.* And that is what makes it hard to accept Hume's results. We can perhaps be led, at least for a moment or two, to suppose that those conclusions might be true of others. But it is hard for any of us to accept the results of the Humean study as true of ourselves. (...) Could I see myself as anything other than a credulous victim of the blind workings of certain 'principles of the imagination'? Is that a satisfactory outcome of the naturalistic study of human beings, including myself?"

One might balk at the claim that a naturalist is committed to the idea that it is only from an external perspective that our thinking and reasoning can be considered in terms of 'oughts'. I concede that a naturalist can say that we qua natural mechanisms differ from a thermostat in that our thought involves the representation of reasons, and that these representations have causal efficacy (i.e., they partly determine combinatory mental processes and their output). But the crucial issue is whether the naturalist is entitled to say that the mental processes that are in part determined by our awareness of 'internal oughts' are causally sensitive to reasons. The point of the computer-analogy I sketched is that on the naturalistic picture, our causal and explanatory thought is analogous to the functioning of a computer in two relevant respects. First, neither we nor the computer have any kind of rational control over the evolution or formation of the patterns or dispositions that cause our respective mental (or quasi-mental) operations (in our case, this includes the representations of 'oughts' and reasons; for instance, our positing of necessary causal connections may be partly determined by a sense of rational appropriateness). Second (relatedly), what kind of patterns of mental operating are developed, is, in both cases, contingent. In the case of the thermostat, its internal structure is contingent upon the intentions of the designer, so there is a sense in which the way the machine works can be considered appropriate from an external perspective. In our case, the development of our patterns to think in causal terms is dependent on processes and conditions that (as I have argued, by appeal to the difficulties that afflict the idea that those processes track corresponding external realities) must be considered non-rational, and that cannot be considered intentional or goal-directed in any deeper sense.

Again, I think that the worries I raised in this section, concerning the incompatibility of the idea that we are rationally justified in our causal thought and the assumption that our causal thoughts "stand under the mechanism of merely subjectively determining [natural] causes", persist in a contemporary framework. Suppose, for instance, we settle (never mind problems concerning preemption or overdetermination) for a Lewisian counterfactual analysis of causation. If the mind is part and parcel of the natural order, then all of the following commitments are the mere upshot of the way our minds happen to have evolved: the very idea of a modal, counterfactual dependence between events; the idea to analyze counterfactuals in terms of talk about similarity and distance of worlds; the 'intuitive' assignment of degrees of similarity and closeness between worlds in particular modal judgments. Alternatively, suppose we settle (never mind problems about the unknowability of abstract entities, the relations between them, and their connectedness to the spatio-temporal world) for the Armstrong-Dretske-Tooley view that causation is a relation of nomic necessitation between universals. If the mind is part and parcel of the natural order, then all of the following commitments are the mere upshot of the way our minds happen to have evolved: the very idea of a necessary connection between events; the very idea of an 'abstract entity'; the very idea that causal necessitation is a relation between universals. None of the ideas that I have enumerated and that are central to the (antagonistic) Lewisian or Dretsikian programs have their origin in sense-perception. The worry I have expounded in this section is that a naturalistic conception of the mind implies that what causes us to think in terms of these ideas are contingent, arbitrary psychological 'habits', whose evolution does (or did) not track objective epistemic reasons. Of course, there might be various evolutionary advantages to thinking in terms of those ideas, but, as we noted earlier, the relation between such advantages and epistemic reasons (rather than pragmatic reasons deriving from survival value) is unclear, especially given the fact that the content of the ideas we are concerned

with does not lend itself to discovery by our mechanisms for detecting features of the world.

I have argued that the idea that our mind is exclusively subject to natural causality suggests that there is no rational justification for the use of concepts, such as causality, that are central to our cognitive orientation and whose content is underdetermined by sense-perception. I do not take myself to have conclusively established this suggestion, because the relevant issues are complex and intractable; and I surely have not been able to exhaust the space of possible naturalistic responses. But I do take myself to have shown, at least, that Kant's worry here is genuine and worth taking seriously.

I now want to suggest that there is a second, conceptually separate problem with the notion that the processes which lead us to think in causal terms are exclusively governed by natural causality. I have repeatedly made use of slightly awkward claims such as, 'Our causal thought is determined by causal laws that are contingent and that might (if the course of nature had gone differently) not have structured the workings of the human mind or brain'. Call this claim P. Now, there would be no problem with the idea that is being expressed in P if we could confidently represent causality as independent of our mind. For then we could say that the world is inherently structured in causal terms, and that if the causal laws that govern the course of nature had turned out differently, the human mind – which, in this scenario, is wholly and solely part of nature – might have failed to apprehend this fundamental constituent of reality. But, as we saw, considering causality (understood as a 'modally thick' feature) as part of mind-independent reality renders causality unknowable and mysterious. If, on the other hand, we suppose that causality is (in part) a mind-dependent feature, then the idea expressed in P seems illegitimate. It states or implies *both* a metaphysical dependence of our mental activity on causality/law-likeness *and* a metaphysical dependence of causality/law-likeness on our mental activity. One can hardly respond that the relation between our mind's activity and the property of natural causality/law-likeness is one of mutual dependency, because an activity cannot produce or impose a property while already exhibiting or instantiating that property. There is a related problem here. In P, we purport to represent a possible world in which the human mind *exhibits a causal structure that prevents it from thinking in causal terms*. Now, the laws of nature are (at least in some sense) contingent – natural reality could have been or become such that events in nature occur according to different laws. If the type of mental act which (according to the 'idealist', anti-realist picture that is invoked to avoid the realist picture that makes causality unknowable and mysterious) contributes the causal relation to reality is *itself* supposed to stand under natural laws, this implies that nature could have been such that our thinking stands under natural causal laws that prevent it from contributing the property of natural, law-like causality to reality. And this is an incoherent notion: if it is assumed that our mental activity is the very source of the property of natural causality/law-likeness, there is a contradiction in the conception that our mental activity might stand under natural causal laws that determine us not to produce the representation and property of causality/law-likeness. The naturalistic picture of the mind is committed to saying that all our mental activity stands under contingent natural laws, and hence it is committed to saying that it must be conceivable that our mental activity might have evolved in such a way that its contingent causal structure prevents it from producing the representation/property of natural causality/law-likeness. Hence, the naturalistic picture cannot accept that the property of natural causality/law-likeness originates in us (in our minds).

A similar point pertains to concepts such as simplicity or unity. The claim that the

property of natural law-likeness depends on our minds, namely, on our commitment to the ideas of simplicity and unity (of scientific theories) is (as we saw; cf. footnote 635) not merely a Kantian claim; it is accepted by contemporary naturalists such as Lewis and Loewer.⁶⁵¹ Now for Lewis and Loewer, the property of being a law is "partly constituted by psychological factors".⁶⁵² This equation of *mind-dependence* with dependence on *contingent psychology* (or neurology) gives rise to the following problem: if the property of law-likeness depends in part on our commitment to simplicity and unity, the notion that this commitment is *itself* a matter of contingent natural law (i.e., a matter of how our mind happened to develop in the natural order of things, with the result being its propensity to think about nature in terms of unity and simplicity) seems incoherent: it *both* presupposes an intrinsic connection between the actual constitution of the mind and the property of nomological regularity *and* asserts the metaphysical independence of the property of nomological regularity from the actual constitution of our mind.

A related point can be made without focusing on the *contingency* of the empirical laws that would govern all of our thinking if 'all causality were mere nature'. If we accept both the mind-dependence of causality and the naturalistic idea that our causal thought is governed by causal laws, it follows that there is a natural law (L) that states that under certain conditions (upon having observed a pattern of regularity, etc.), the human mind will posit the property of natural law-likeness (*mutatis mutandis* for simplicity and unity). Such a law would have to be considered as empty of content. The point and function of an empirical law is to explain the occurrence of events or the emergence of properties by appeal to conditions that are causally sufficient for such occurrence or emergence. The point of (L) would thus be to explain the human mind's tendency to posit the properties of empirical law-likeness and empirical necessity. But if the human mind is the very source of this property, then we cannot invoke this property in trying to explain the mind's contribution of this property. That is, we cannot explain the emergence of empirical law-likeness by appealing to some *law-like* mechanism, i.e., a mechanism that already instantiates the explanandum. This would be like explaining the emergence of some chemical property by appealing to a mechanism that must be described in terms of that very property.⁶⁵³

Here we must also take into account the peculiar epistemological status of the causal principle in Kant's theory. Kant attempts to legitimize the (schematized) concept of causality, and

⁶⁵¹ Loewer 2004, p. 191: "So it is apparently a consequence of Lewis' account that which propositions are laws depends on mental facts about us."

⁶⁵² See Loewer 2004, p. 191. See also footnotes 635 and 637 above.

⁶⁵³ This is the problem with Patricia Kitcher's idea (1999, p. 434 ff.) that "once we realize that the causal relation is itself contributed by the actions of our faculties, then we might wish to [say that]...our data on the actions of our faculties and our data on the elements of our knowledge representations are sufficiently orderly that...we interpret our faculties as causes of some key elements of representation." The 'key element of representation' at issue here *is* the causal relation. But a *causal law* purporting to explain how our faculties produce the *causal relation* must be regarded as empty, for it presupposes the explanandum. As I argued earlier (chapter X, footnote 308), it does not suffice here for Kitcher to reply that this "reapplication of the model to itself" only shows that "Kant's epistemological theory has...the same status as a scientific theory". For sure, any ordinary scientific theory that attempts to arrive at empirical laws will (quite rightly so) take for granted that there is such a thing as empirical law-likeness. But the scientific theory that Kitcher envisages could not take this for granted: it could not simply attempt to uncover a mechanism instantiating the property of law-likeness. This is why Kant himself thought of his account of causality as prior to and independent of scientific theorizing. On any broadly Kantian account of causality the act of contributing the causal relation to empirical reality has metaphysical priority over the causal structure of reality, and hence this act cannot coherently be represented as depending on that structure.

hence the causal principle, by arguing that applying this concept and principle is a condition for the possibility of objective thought: its validity hinges on the idea that in order to represent a natural alteration from state s_1 to s_2 , one must refer the alteration to some preceding event from which it follows invariably and necessarily; for it is only thus, Kant argues, that the original state s_1 can be represented as objectively temporally prior to the resultant state s_2 . Now, if the formation of the concept and principle of causality is supposed to be a natural event and hence a potential object of empirical observation and explanation, the attempt to observe and explain this act requires that one *makes use of* the concept and principle of causality, namely, that one represents the formation of this concept/principle as following invariably and necessarily from some preceding event (a representation which is the content of the concept/principle of causality). Thus, the concept and principle of causality is epistemically prior to any attempt to explain or even objectively represent the processes that are supposed to give rise to this concept/principle. This implies that any attempt to scientifically understand the representation of causality (via uncovering the processes from which it derives) must already invoke and rely on this very representation. This (again) seems to frustrate the characteristic scientific goal to get a grip on an explanandum (here, a certain kind of conceptual representation) without depending, for one's explanation (i.e., for characterizing, even for objectively representing the processes that are meant to constitute the explanans), on what is to be explained. I believe that this general point has a great deal of plausibility independently of the specific way in which Kant tries to make it (i.e., independently of the idea that an appeal to causality is necessary for objective time-determination). The concept of causality (simplicity, unity...) is so fundamental to our cognitive efforts that no serious attempt at cognition, understanding, or explanation can dispense with this concept: any attempt to understand a phenomenon must make use of the notion of a sufficient explanatory reason and hence of the notion of a causal condition. If it is a constraint on a satisfactory explanation that it does not invoke the explanandum – a point which, admittedly, I have presupposed here – then there can be no satisfactory explanation of how the mind comes up with the representation of causality.

The common source of all these problems is the attempt to reap the benefits of the Kantian defense of causal necessity without investing the price required by the Kantian enterprise. The benefit of the Kantian account of the causal connectedness or law-likeness of natural events is that it traces the source of these properties to the human mind, namely, to what it takes for the human mind to have cognition of objects. This might suffice to answer the epistemological worry raised by Hume, namely, the worry that there is no way to legitimize the claim that natural events are causally connected either on the basis of empirical observation or on the basis of logical, conceptual analysis. The price this account requires one to pay is the willingness to forsake, first, the notion that we have insight into the absolute structure of metaphysical reality as it is independently of the human mind, and, second, the concomitant notion that all reality as such is governed by natural causality. An account (such as Patricia Kitcher's) which holds that the processes which lead us to 'contribute' the property of natural causality to reality are themselves governed by natural causality attempts to have it both ways: it purports, in an idealist fashion, to trace the property of natural causality to the human mind, and it also pretends, in a realist/naturalist fashion, that everything within reality can be comprehended in terms of natural causality. But, as I tried to show, this combination of Kantian idealism and anti-Kantian naturalism seems incoherent.

5: Concluding remarks

In the preceding section, I have tried to uncover what I take to be the fundamental root of Kant's conception of the 'pure spontaneity of the understanding' (or, if we focus on ideas such as unity or simplicity, of the pure spontaneity of theoretical reason).⁶⁵⁴ I have argued that Kant's appeal to pure epistemic spontaneity rests on the twofold idea that (a) on a realist conception of features corresponding to 'pure' concepts, such as causality, that seem central to our cognitive orientation, these features are rendered unknowable and mysterious and (b) that an idealist conception of (the origin of) features such as causality is incompatible with an exclusively naturalistic conception of the human mind. Hence, if one thinks that a commitment to a kind of mental activity that is not governed by natural causality is overly suspicious, one must forgo the Kantian idealist attempt of defending the legitimacy of the concept of causality, and one must seek to legitimize this concept by appeal to the absolute, mind-independent structure of reality. I have certainly not shown that this is an impossible task. But in the face of Hume's epistemological challenge, it seems at least reasonable for Kant to suspect that the task cannot be completed.

This concludes the argument of this chapter. We started with the idea that worries about a naturalistic conception of empirical belief-formation can be defused by appeal to the fact that our thinking about nature is causally constrained by those very natural objects and events toward which our empirical thinking is intentionally directed. I suggested various ways in which the appeal to a causal constraint that the external world imposes on our empirical beliefs cannot preempt worries about the rational justifiability of those beliefs or of the conceptual schemes employed in those beliefs under the assumption that our beliefs and concepts have their origin in contingent psychological tendencies (or in underlying contingent neurological mechanisms).

This shows that one cannot, at least not easily or uncontroversially, suppose that autonomy from nature is an ideal of practical reason alone that our theoretical (causal-explanatory) perspective on the world exposes as confused or illegitimate. For without supposing that the naturalistic standpoint exhibits some kind of autonomy from nature in forming and applying the concepts (such as causality) that figure in its claims, the objectivity of the naturalistic explanations that it purports to offer might, upon closer examination, elude us.

⁶⁵⁴ In uncovering this root, I have tried to answer Ameriks' challenge to Kant's incompatibilism about theoretical spontaneity. Ameriks suggests (2000, p. 207) "that although what we intend when we use our understanding and reason may appear transcendentally intended, our intending it none the less might be *sufficiently explained* by a natural process that involves nothing transcendent in us". For the reasons I have tried to spell out, I think that Kant has reasonable grounds for rejecting the supposition that spontaneous acts of forming and using the pure concepts of the understanding and the pure ideas of theoretical reason are exclusively governed by natural laws.

Chapter XIII: Morality as the Ratio Cognoscendi of Practical Freedom

In the preceding chapters, I have been primarily concerned with Kant's conception of the kind of freedom and spontaneity involved in empirical cognition. In this chapter, I want to argue that the appeal to the spontaneity of our higher *epistemic* faculties plays a crucial role in Kant's defense of the spontaneity of our higher *practical* faculties: Wille (as practical reason) and Willkür.

1: Two dimensions of Kant's justification of the belief in free will

The problem of free will looms large in the Critique of Pure Reason, in the Groundwork, and in the Critique of Practical Reason. There is a problem for Kant because his concept of practical freedom is (as I argued in chapters II and III) metaphysically ambitious: to ascribe freedom of will is to ascribe causal capacities whose exercise does not stand under conditions of time and deterministic laws of nature. What could possibly justify the claim that we have such non-natural causal capacities, whose exercise cannot possibly be an object of sensible intuition?

To understand the logic and force as well as the difficulties that afflict Kant's justification of the idea of free will, we must first acknowledge that Kant's project of justifying this idea is complex and has a number of distinct (albeit closely related) dimensions. The *first* and most basic layer of justification is provided by Kant's claim that conditions of space, time and nature are not metaphysically fundamental: space and time are only forms of human sensibility, and the principle that every event is determined by another in time according to necessitating laws holds true only with regards to the empirical character of objects that ontologically depends on the character of our cognitive faculties. This is the doctrine of transcendental idealism, which is established by theoretical reason on the basis of general epistemological considerations. Since temporality and natural (deterministic) causality are not metaphysically fundamental (i.e., do not constitute absolute metaphysical conditions on anything that may exist or happen) within the framework of transcendental idealism, this idealism opens up ontological space for the existence of the (atemporal, non-deterministic) causality of freedom; and it thereby opens up logical space for combining the belief that we (considered as things in themselves) are free with the belief that we (considered as appearances) are causally determined. (See my chapters I and II for a detailed explanation of the metaphysical, idealist framework of Kant's theory of freedom).

The *second* dimension of the justification of the belief that we are practically free relates to Kant's standpoint distinction. Because the idea of freedom designates an unconditioned, atemporal causality, experience cannot provide us with evidence for the actual instantiation of this idea, and – worse – this idea even violates the conditions of experience (such as temporality). This is why we cannot use this idea in the context of trying to observe and understand the natural world: it cannot be invoked to give us theoretical knowledge of what objects are like, of why people act as they do, or of what happens when someone acts. In other words, we *must not* employ the idea of freedom to characterize objects from the empirical standpoint. So in order for this idea to have a legitimate employment, it must be possible to specify a point of view from which the standard for adopting beliefs is not fixed by empirical evidence (experiential data), and from which the purpose of using concepts is not that of understanding (observing, explaining, and predicting) what happens. This standpoint is that from which we aim at establishing prescriptive normative principles that specify how one ought to act. Since we do not use the concept of freedom for the purposes of achieving objective experience of nature, it is not an objection to our use of that concept that it cannot be instantiated in experience

and that it violates the conditions of objective experience. (See my chapter VIII for a detailed discussion of Kant's standpoint distinction.)

Important as they are, these two dimensions cannot *exhaust* Kant's attempt to legitimize the concept of practical freedom. We can see this by contemplating the credentials of a further non-empirical concept, the concept of the power of foresight. It is conceivable that there might be such a power, although our scientific theories give us every reason to think that nature could not contain such powers (CPR, A 223/B 270). This concept satisfies the conditions suggested by the two dimensions of justification considered so far. If Kant's idealism creates ontological space for non-natural properties, this space might be occupied by powers that the known laws of nature rule out. Moreover, we can delineate a non-empirical *standpoint* for the use of this concept. Consider a 'religious standpoint' that relinquishes all explanatory pretensions and is concerned only with questions of faith. This standpoint might require the belief that the prophet's power of foresight exists. Since someone occupying this standpoint does not use the concept of foresight for the purposes of cognition, it is not an objection to her use of this concept that it cannot be given in experience and conflicts with our scientific conception of what is (really) possible.

Kant does *not* claim that we have reason to believe in the power of foresight. What this example suggests is that his appeal to transcendental idealism and his appeal to a distinction between standpoints are necessary but insufficient for legitimizing the belief in freedom.

2: Morality and free will: the threat of a circle and the 'fact of reason'

What sets apart the belief in freedom of will from the belief in foresight is that the former belief is integral to our moral practices: our conviction that we have free will underlies our conviction that we are both authors of and subject to categorical imperatives, and that we can consider ourselves responsible for both complying with and for violating those imperatives.⁶⁵⁵ Since our self-conception as moral agents is (unlike the belief in prophets and foresight) an essential, inescapable aspect of our self-conscious awareness and rational orientation, it may serve as a non-arbitrary basis for establishing that we have the kind of freedom that seems required by this self-conception. Relatedly, if moral oughts qualify (unlike the belief in foresight) as intersubjectively binding and thereby as objective cognitions (of duty), we can attribute genuine legitimizing force to the normative standpoint from which one simultaneously recognizes these principles and one's practical freedom (as opposed to the 'religious standpoint' that seeks to infer to the existence of foresight on the basis of obscure mystical thoughts about prophecy).

Kant posits a close, indeed a reciprocal relation between the concepts of morality and freedom of will.⁶⁵⁶ But *precisely because* free will and morality are so closely tied to one another, it is hard to see how we could justify one of them *via* the other. If we were justified in believing that we have free will, we would be entitled to infer that we can produce and comply with necessary moral laws. If we were justified in believing that that we can produce and comply with moral laws, we would be entitled to infer we are free. But this are mere bi-conditionals whose truth leaves open whether we *really are* free or whether our moral practice (as centered around the conviction that we are bound by valid moral laws) *really is* legitimate. I think that this point

⁶⁵⁵ See part 2 (chapters IV, V, VI and VII) for arguments that support these claims.

⁶⁵⁶ This holds only for the 'positive' concept of freedom. The 'reciprocity thesis' says that the only positive conception of practically intelligible, lawful free action is that of action on the basis of self-imposed moral laws. It does not mean that a will who chooses to act contrary to moral reasons is not free in the 'negative' sense that signifies absence of determination by natural causes. See my chapter III, section 3.

is *at least part of* the source of Kant's worry about the threat of a circle in his argument in the Groundwork.⁶⁵⁷ If the validity of the ideas of moral obligation and moral responsibility implicates our freedom, the bars for legitimizing the ideas of moral obligation and responsibility must be as high as the bars for legitimizing the belief in freedom *qua* necessary condition of the possibility of obligation and responsibility. Given the metaphysical ambitiousness of Kant's concept of freedom, the bars for legitimizing a belief in such freedom are very high; doubts about whether we are capable of reasoning, judging and choosing independently of determination by natural causes are easy to motivate. In other words, if our conception of morality and responsibility requires for its legitimacy an ambitious concept of freedom, we cannot simply take *its* legitimacy for granted and side-step questions about what justifies our

⁶⁵⁷ See GMS, 4: 448-50. I am not claiming here to offer an analysis of all possible facets of Kant's worry about a circularity in his argument in the Groundwork. That would require a paper, maybe a book on its own. All I am claiming is that the line of thought I sketched above partly drives Kant's worry. In the remainder of this footnote I want to defend this interpretation against some prominent or representative challenges. Paton (1958, p. 225) famously claims that Kant's worry about circularity misrepresents his actual argument since "he...professed...to establish the presupposition of freedom by an insight into the nature of self-conscious reason quite independently of moral considerations". As should be clear from my argument in the preceding chapters, I agree that Kant considers reason as such, in both its theoretical and its practical function, free. But it does not follow that moral considerations are irrelevant to Kant's appeal to the nature of self-conscious reason as such. When Kant makes this appeal (GMS, 4: 448), he invokes our self-conception as subjects whose practical reason is *pure* and can establish practical *laws*: this implies moral considerations, because *pure* practical reason would, if it did exist, establish *unconditionally* binding laws, and these would be categorical imperatives. Kant's worry about the notion that morality might be a mere 'chimerical idea' (GMS, 4: 445) concerns in part the possibility that practical reason might always be conditioned, in its judgments, by a contingent natural ground so that there would be no categorical imperatives (see GMS, 4: 419, and chapter V). I also disagree with Ameriks' claim (2000, pp. 203-6) that Kant's real worry is not about the circularity inherent in the move from morality to 'negative' freedom (i.e., the absence of causal determination), but about the move from negative freedom to positive freedom (morality). If it could be shown that practical reason is negatively free, this would amount to a proof of the idea that practical reason can operate without being conditioned by contingent natural grounds, and the principles it could produce under this premise would (in virtue of their independence from contingent desires) qualify as *laws* which as such must be the imperatives of duty. The idea that practical laws which are independent of all contingent empirical conditions must be the moral laws of duty (which derive from the categorical imperative) is the upshot of what Kant calls (GMS, 4: 445) the analytic argument enunciated in section II of the Groundwork. *If* reason itself rather than nature gives the law (GMS, 4: 444), i.e., *if* the origin of practical rules is autonomy rather than heteronomy, then the law must be the categorical imperative. But *that* reason is capable of self-legislation independently of determination by natural causes is not something that Kant asserts prior to section III of the Groundwork. 'Negative freedom' is a metaphysical condition of positive freedom (i.e., of autonomy, which in the practical sphere is morality). Since I think that there is a clear sense in which Kant's *own* position is potentially afflicted by the threat of a circle, I am not convinced by Brandt's intriguing 1988 idea that the circle refers to the Wolffian school. Admittedly, my interpretation takes Kant to be concerned, in part, with our freedom of will *qua* practical reason. Allison claims (1990, p. 226) that in the Groundwork Kant does not yet "possess" the distinction between will *qua* practical reason and Willkür. It is true that Kant explicitly draws this distinction only in the Metaphysics of Morals, but he surely is aware of the fact that we (take ourselves to) have a distinctive capacity for reasoning toward practical laws already in the Groundwork. Allison (whose account is, in this respect, similar to Paton's) thinks (ibid., pp. 216) that (at GMS, 4: 448) Kant's appeal to our self-awareness as authors of pure principles pertains only to "reason in its cognitive activity (judgment)", and that an independent step is required to get from there to reason "in its practical activity...as will". But in the relevant passage, Kant proceeds from the idea that reason must regard itself as the author of its own principles to the claim that it must "consequently" regard itself as free "*qua* practical reason, or *qua* will of a rational being". So the appeal to the freedom of reason *qua* author of *practical* judgments (that represent laws of action) is very explicit in the text.

belief in its metaphysical foundations. Unless we have reason to assume that these foundations exist, we cannot presuppose that our attachment to morality *is rational*. Since our attachment to the idea of morality is based on our attachment to the idea of freedom, grounds for suspicion about whether the latter idea has objective validity give rise to a worry that the former idea is likewise "a chimerical idea without truth" or a mere "Hirngespinnst" (GMS, 4: 445).

Thus, it seems that certain extra-moral grounds are needed for a non-circular justification of the concepts of free will and morality. If an appeal to our moral self-conception is to serve as the foundation of defending the legitimacy of our self-image as practically free agents, we must have some assurance for the validity of our moral self-conception that does not already presuppose either the validity of our moral self-conception or the legitimacy of its necessary condition, the metaphysically ambitious idea of free will.

Now, some commentators concede that Kant, *prior to* the Critique of Practical Reason, tried to legitimize the ideas of morality and free will by appeal to the kind of freedom exhibited by our epistemic faculties; they hold that Kant here relied on an *inference* from theoretical to practical freedom that, as he came to realize, is a complete failure.⁶⁵⁸ The overwhelming consensus is that an appeal to extra-moral considerations is completely banned in the Critique of Practical Reason, where the appeal to the moral law as a 'fact of reason' is meant to stand on its own.⁶⁵⁹ Commentators differ in their assessment of the success of this strategy: for some, the appeal to the moral law as a fact of reason marks "the epistemic point of culmination of Kant's theory of freedom"⁶⁶⁰, whereas for others it signifies a lapse into rationalistic dogmatism.⁶⁶¹

I side here with the critics: I do not think that Kant's appeal to the consciousness of the moral law as a fact of reason⁶⁶² can *all by itself* serve to legitimize our idea of practical freedom.

⁶⁵⁸ See Allison 1990, p. 69, pp. 227-8; Henrich 1973, p. 245 ff. and 1975, p. 72, and (following Henrich) Schönecker 1999, p. 209 ff. and p. 299 ff.

⁶⁵⁹ Those who believe that Kant gave up the attempt of a deduction of morality and freedom from non-moral premises after he realized the failure of an attempt at such a deduction in the Groundwork include Allison 1990, chapter 12; Ameriks 2000, chapter 6; Ilting 1972, p. 125; Beck 1960, p. 165 (but note that Beck still seems to find an attempt toward a deduction of morality in the Second Critique); Rawls 1989 and Willaschek 1992, p. 171 ff. Henrich (1973, 1975) argues that this change of strategy came prior to the Groundwork and that the *Groundwork* already contains, in essence, the same strategy as the Second Critique (compare Paton 1958). This is related to the view, defended by Brandt 1988, Freudiger 1993 and Reibenschuh 1990, that when Kant speaks of a 'deduction' in the Groundwork, this means something different than in the First Critique, and that the arguments of the *Groundwork* and the Second Critique can be considered complementary once this aberrant sense of 'deduction' is taken into account. Bojanowski (2006, pp. 59-60 and pp. 209-28) also suggests that the notion of a deduction in the *Groundwork* is a modest, 'analytic' one, but argues that the approach of the Second Critique is nonetheless wholly different and renders the argument of the Groundwork (not false but) superfluous.

⁶⁶⁰ Bojanowski 2006, p. 191 (my translation). See also Allison 1990, chapter 13 and Henrich 1973.

⁶⁶¹ For this diagnosis, see Ameriks 2000, p. 211 ff.; Bittner 1983, pp. 138-45; and Prauss 1983, pp. 62-70 ff..

⁶⁶² There is a question here about whether we should take the fact of reason to refer to our consciousness of the moral law (Allison 1990, p. 233) or to the fact that reason gives such a law (Beck 1960, pp. 200-14; Bojanowski 2006, pp. 61-2; Willaschek 1991, 1992). Allison argues that if Kant understood the 'fact' in the latter sense, his argument would be question-begging; Willaschek gives a very illuminating analysis of Kant's general notion of a Faktum *qua* legislative deed (Tat). I suggest a compromise here: while Allison is right that Kant's argument must proceed from our conscious awareness of the moral law, one can construe the object of this awareness as the moral law *qua* product of a legislative deed. This is not question-begging because Kant can still be taken to address the question of whether we have reason to consider this consciousness veridical. Against Allison (*ibid.*, p. 233) and T.C. Williams (1968, pp. 103-11), I do not think that we need to drive a wedge here between the awareness of the moral law as a formal principle and the particular moral rules deriving from it, because an

To see why not, let us look at a passage in the Groundwork where Kant himself raises doubts about whether we can take our consciousness of unconditional moral demands at face value:

...the question how the imperative of morality is possible, is undoubtedly one, the only one, demanding a solution, as...the objective necessity which it presents cannot rest on any hypothesis, as is the case with the hypothetical imperatives. Only here we must never leave out of consideration that we cannot make out by any example, in other words empirically, whether there is such an imperative at all, but it is rather to be feared that all those which seem to be categorical may yet be at bottom hypothetical. (...) Who can prove by experience the non-existence of a cause when all that experience tells us is that we do not perceive it? But in such a case the so-called moral imperative, which as such appears to be categorical and unconditional, would in reality be only a pragmatic precept, drawing our attention to our own interests and merely teaching us to take these into consideration. (GMS, 4: 419)

The worry that is being expressed here is – as is often the case in Kant – at once metaphysical and epistemological. The concept of a pure will *qua* practical reason that produces necessary prescriptions in complete independence of the influence of sensible causes (an influence that would negate the claim to objective necessity inherent in self-proclaimed practical laws) is metaphysically demanding; and since such a will must be considered a non-empirical capacity, its activity cannot be an object of perception. But then how could we ever claim to *know that* our practical reason is capable of a purely spontaneous exercise? The mere fact that we do not perceive the influence of sensible causes on our reasoning, judging and choice does not prove that there is no such cause that operates secretly and escapes our impoverished capacities for tracing the influence of psychological causes. The same goes for any putative *defense* of the necessity of the moral law: any reflection that purports to show that the moral law has unconditional authority would be illusory if it derived its influence on us from some contingent sensible motive; for instance, a defense of the idea that we *must* treat rational beings as ends in themselves would be deceptive if its appeal for us did depend on our contingent empirical constitution. If we cannot know that reason's activity is pure, we cannot be confident either that our awareness of practical necessity is veridical. The veridicality of our consciousness of the moral law implicates the veridicality of our consciousness of pure, autonomous practical reason.

I want to forestall an objection to my construal of the dialectical situation. I have been suggesting that a significant part of the problem that Kant faces here is that in order to consider our awareness of practical laws veridical, we must presuppose the freedom of our practical reason from determination by natural causes. One might object that this involves a double misunderstanding of Kant's concerns, for his concern about moral principles relates only to their *content* rather than to their *causes*, and his concern about the metaphysics of free agency relates only to our capacity to act in accordance with moral laws. As Guyer puts it:

The transition to a critique of pure practical reason in Groundwork III concerns not the content of the principle of morality but the fact of our freedom to act in accordance with the canon of rationality which that principle expresses. (...) ...a metaphysics of morals is metaphysical because it shows a moral principle to be a priori rather than empirical, not because it adduces a metaphysical conception of the moral agent.⁶⁶³

I have addressed variations of this objection in previous chapters⁶⁶⁴; here I shall simply

awareness of such rules involves, essentially, an awareness of the moral ideas contained in the form of law.

⁶⁶³ Guyer 1995, p. 358 (footnote 5). See also Beck 1960, p. 172, 185 and Bojanowski 2006, p. 242-4, p. 261.

⁶⁶⁴ See, in particular, chapter III, section 3 (especially footnote 151); and the overall argument of chapter V.

recapitulate a few relevant points. I concede that Kant employs a (narrow) sense of 'spontaneous causality' that applies to acts that are (or fail to be!) in accordance with principles of morality. But he also has a wider (or 'generic') notion of free, spontaneous agency that applies to practical reasoning and judging; he uses this wider notion when he refers (CPR, A 803/B 831) to "...reason in these actions, through which it prescribes laws..."; and he considers the act of practical legislation free in a sense that *precisely* fits his metaphysically loaded notion of freedom from determination by natural causes.⁶⁶⁵ Moreover, we should not drive a wedge between the causes and the content of practical representations: for Kant, the a priori content of a representation is tied to its a priori origin. He states in very explicit terms that the supposition that a practical judgment is caused by contingent empirical causes is incompatible with the supposition that the judgment has a priori, necessary validity: in an anticipation of a Humean 'desire in, desire out' conception of practical reasoning, he writes that a contingent desire can only produce (hervorbringen) a want but not a necessary 'ought' (CPR, A 548/B 576). As I argued in chapter V, this is not arbitrary: it is rather hard to see what point there is in insisting on "the independence of the principle of morality from any contingent and empirically given end"⁶⁶⁶ *with respect to its content* if, at the same time, one happily concedes that our reasoning toward (and 'legislating') this principle is causally dependent on some contingent, empirically given desire.

To be clear: I accept that Kant – at least in the Groundwork and the Second Critique⁶⁶⁷ – posits a tight, inseparable link between the idea that practical reason can all by itself legislate principles and the idea that practical reason can all by itself move us to act in accordance with these principles. But *precisely because* of this link, one can raise metaphysical (and related epistemological) worries not only about the idea that our choices can be based on pure (desire-independent) motivation, but *also and in the first place* about the idea that we are capable of pure acts of practical judgment that are independent of the causal influence of *contingent* desires and (hence) that our consciousness of practical *necessity* is veridical (this point is expounded in somewhat greater length in chapter V, section 1). I shall return to the implications of this link between purity of practical reason and purity of Willkür in section 6 below.

For present purposes, the important question is: given the metaphysical ambitiousness of the idea that we can reason and judge independently of the influence of contingent sensible causes, how is Kant's appeal to the 'fact of reason' supposed to legitimize this idea? Those who are enthusiastic about Kant's putative strategy of relying *solely* on the moral law as a fact of

⁶⁶⁵ In acts of practical legislation, we presuppose that our reason is not "receiving a bias from any other quarter with respect to its judgements, for then the subject would ascribe the determination of its judgement not to its own reason, but to an impulse. It must regard itself as the author of its principles independent of foreign influences" (GMS, 4: 448). Independence from determination by sensuous impulses signifies the absolute spontaneity of transcendental freedom.

⁶⁶⁶ Guyer 1995, p. 369. I find of much value in Guyer's discussion of the various different questions that one can raise about the possibility of a categorical imperative. But I am puzzled by his idea (p. 384) that the conditions for the "real possibility" of the categorical imperative are demonstrated, in some sense, prior to addressing any metaphysical questions about the capacities required for *both* acts of practical legislation *and* acts of compliance with the relevant laws (i.e., prior to section III of the Groundwork). If, as Kant concedes, it is still (prior to Groundwork III) an open question whether the categorical imperative is an empty 'Hirngespinnst', then it seems to me that we cannot yet assume the real possibility (in any relevant sense) of the categorical imperative either.

⁶⁶⁷ Kant seems to take for granted our capacity to produce necessary practical prescriptions, while questioning our capacity for a priori motivation, in the Critique of Pure Reason, whose conception of moral motivation is significantly at odds with that of the Groundwork and the Second Critique. See my chapter III, section 2.

reason must think that this fact suffices to disarm the metaphysical and epistemic doubts that Kant raises in the passage quoted above. How could this be? Henry Allison responds:

The main point is simply that if [the consciousness of freedom] were illusory, then our autonomy or positive freedom and with it the whole conception of ourselves as moral agents would be also. But this has been shown to be actual by the fact of reason. Consequently, if the fact of reason is, indeed, the fact of reason, then this consciousness cannot be regarded as illusory.⁶⁶⁸

For Allison the fact of reason is our consciousness of standing under moral laws.⁶⁶⁹ Hence, he seems to be arguing here as follows: (I) The veridicality of our self-awareness as *free* agents is a necessary condition of the veridicality of our self-conception as *moral* agents. (II) Our self-awareness as *moral* agents has been vindicated as veridical by our consciousness of standing under moral laws (i.e., "has been shown to be actual by the fact of reason"). (III) Hence, our self-awareness as (positively) *free* agents is also vindicated as veridical.

But it seems to me that this argument is viciously circular. For in *simply claiming* (II) that we can indeed take our consciousness as standing under moral laws at face value, we are *already* presupposing (III) our freedom (as a necessary condition of the moral law) without even having addressed the metaphysical and epistemic worries that arise here (and that Kant himself raises in the Groundwork passage considered above). Allison is correct to emphasize that we cannot consider our consciousness of our freedom as illusory *insofar as* we assert our status as moral agents, but this is nothing but a reminder of the reciprocal relation between morality and freedom. It suggests no way out of the circle that Kant was rightfully concerned about in the Groundwork. Unless we have a way to answer the doubts that afflict the claim that we can take our consciousness of the moral law at face value, we cannot infer from this consciousness to our freedom (either in its negative or its positive sense).

Defenders of the claim that the appeal to our awareness of the moral law is sufficient to justify the claim that this awareness is veridical might respond in various ways. First, they might hold that Kant here presupposes only his analysis of our common moral concepts, namely, the fact that we *do* draw a difference between counsels of prudence and moral demands and that we *do* abstract in our moral reasoning from contingent empirical interests; hence, he does not need to further *establish that* moral judgments make categorical demands.⁶⁷⁰ I agree that Kant presupposes, rightfully so, that our moral practice is not a system of hypothetical imperatives. But the mere fact that we *commonly take* our moral judgments to be categorical imperatives cannot preempt the metaphysical and epistemic doubts about what this awareness presupposes for its veridicality. The metaphysical doubt pertains to the idea that we have a non-natural capacity of practical reason that operates independently of conditioning sensible grounds (e.g., desires); the epistemic doubt pertains to the fact that we cannot intuit the actual workings of this capacity. These doubts cannot be quelled merely by appeal to what we commonly and unreflectively presuppose. Unless we can find some kind of support for the metaphysically demanding idea that the exercise of our capacities for reasoning and judgment is not simply a

⁶⁶⁸ Allison 1990, p. 247.

⁶⁶⁹ Allison 1990, p. 233, p. 236.

⁶⁷⁰ See Bojanowski 2006, pp. 68-70, although it is not quite clear to me whether he takes this point to provide a defense of the moral law against (1) metaphysical and epistemological or rather against (2) normative challenges (for this distinction, see section 6 below). If he has only (2) in mind, I agree that the appeal to what is implicit in our ordinary practices is intended by Kant to be sufficient.

function of our contingent psychological constitution (the Humean 'desire in, desire out' conception of practical reason), we cannot confidently rule out the possibility that the principles that present themselves to us as categorical practical laws are (as Kant puts it in the above Groundwork passage) nothing but concealed hypothetical imperatives.

A somewhat different defense of the claim that the appeal to the moral law can stand on its own appeals to the idea that given the great significance that morality plays in our life, there are *conclusive practical reasons that entitle us to believe* in the veridicality of our moral consciousness, and hence in freedom.⁶⁷¹ But this idea raises two problems. Doubts about whether we have grounds for believing in the existence of freedom qua metaphysical foundation of morality imply doubts about the veridicality of our moral consciousness. So unless these doubts can be quelled, the appeal to a *practical need* for the belief in freedom that is rooted in an attachment to morality might reduce to the appeal to a strong *desire* to believe in our freedom that is rooted in a deeply seated socio-psychological attachment to morality. And while this might give us practical, desire-based reasons for the belief in freedom, these are the wrong kinds of reasons for *legitimizing* this belief. Consider again the belief in foresight. We can easily imagine circumstances in which someone has a strong desire to believe, on the basis of her attachment to the Bible and its tales of prophets, that people have (had) this capacity. But this does not render either the belief in foresight or the underlying attachment to the Bible rationally justified. More generally, if one allows that an appeal to practical needs suffices to justify the bracketing of evidential standards for belief, one opens the floodgates for wishful thinking. This brings us to the second problem: the notion that we can adopt a belief *solely* on the basis of practical needs implies a kind of voluntarism that is not only questionable in itself but also (as we saw in chapter IX) explicitly rejected by Kant. It seems that whatever attitude we can adopt willfully on the basis of practical needs, in suspension of evidential standards, cannot be belief (Kant's 'assent'). Of course it is possible to adopt beliefs unconsciously through desire-based processes. But those beliefs can hardly be considered rational or well-grounded.

A third option for the proponent of the claim that Kant's appeal to the fact of reason is meant to be self-legitimizing is to argue that doubts about the existence of the metaphysical foundations of morality can be quelled, and our moral consciousness can be validated, by appeal to the fact that this consciousness impinges on us *only from a practical point of view*. This idea might be understood in terms of the second dimension of justification explained above: i.e., in terms of the idea that from a standpoint where we try to establish what ought to be done, we are not bound by the evidential standards that govern beliefs about the properties of natural objects. But this is a purely negative point that emphasizes only that the evidential standard that *would* license the belief that our moral consciousness is veridical is *not* the same as that which licenses beliefs we adopt in contexts of theorizing about nature. It does not specify what evidence *does* license the belief in the validity of our moral consciousness in view of the metaphysical and epistemic challenges this belief faces. Unless some such evidence is specified, the belief continues to look rationally suspicious, *even from a practical point of view*.⁶⁷²

⁶⁷¹ If I understand him correctly, Pereboom 2005 assumes that Kant tries to support the belief in freedom by appeal to practical reasons. This might also be Korsgaard's point when she suggests that "we must believe these things [i.e. that we are free] in order to obey the categorical imperative" (1996, pp. 175-6). However, as we noticed before (see chapter II, section 2; chapter VIII, section 7 and footnote 673 below) it is far from clear whether Korsgaard is really concerned with the attitude of *belief* at all.

⁶⁷² In the Canon of Pure Reason, Kant claims that in the practical context in which we figure out what to do,

There is, however, a second version of the appeal to the practical standpoint which holds that the fixation on *evidential reasons for belief* displayed by my argument so far is misguided: from this standpoint we do not aim at adopting theoretical attitudes such as belief at all. Rather, the point of adopting this standpoint is that we take a *practical interest* morality that allows us to make morally good *choices*.⁶⁷³ This suggestion faces two objections, one of which I have expounded before (cf. chapter II, section 3): an interpretation that represents Kant as abjuring the belief in our freedom trivializes his laboring toward a justification of the concept of freedom and falsifies his claim that legitimizing this concept requires his idealism.⁶⁷⁴ Apart from this interpretive worry, the suggestion that aiming at a pure interest in morality preempts doubts about the veridicality of our moral consciousness is too evasive. The kind of interest we morally ought to have in the moral law derives from our awareness of the objective, necessary validity and authority of the moral law. So our practical interest in the moral law is conceptually and causally posterior to our consciousness of moral necessity. But then the appeal to a 'pure' interest formed in response to a veridical *recognition* of the unconditional validity of the moral law *already presupposes* that our awareness of the necessity of moral prescriptions is veridical and hence that *both* our practical reason, qua source of moral prescriptions, *and* our interest in moral prescriptions are independent of contingent empirical grounds. So the notion that we can take a sensibly unconditioned interest in morality is afflicted by the same doubts as the idea that we can make necessary practical prescriptions, and hence it cannot rationally be taken for granted.

I conclude that the critics of the assumption that an appeal to the moral law can *all by itself* legitimize freedom and morality are correct. If Kant *did* rest his case solely on our 'naïve' practical self-consciousness, his justification of morality and freedom would have to be considered unacceptably dogmatic.

3: Theoretical and practical freedom: preliminary considerations

I do not think that Kant intends his appeal to the moral law as a fact of reason to legitimize freedom and morality all by itself. In what follows, I will argue that this appeal is, even in the Critique of Practical Reason, crucially informed and supported by an appeal to theoretical spontaneity. I want to begin by laying out in some detail how I conceive of this strategy. Here we can notice that there are two conditions that must be satisfied in order for the appeal to the

'speculative' concerns about whether we are transcendently free can be bracketed. Many argue that his position here is based on a pre-critical, compatibilist conception of practical freedom. Even if this is incorrect (cf. my chapter III), the positions of the Canon and the Second Critique crucially differ insofar as in the latter work the concern with the *transcendental* sense of practical freedom is in the foreground of discussion.

⁶⁷³ See Allison 1990, p. 57: "Kant is...claiming merely that it is necessary to appeal to the transcendental idea of freedom in order to conceive of ourselves as rational (practically free) agents, not that we must actually be free in the transcendental sense in order to be free in the practical sense." This requires "no ontological conclusions regarding the absolute spontaneity of the self" (1996, p. 54); rather (1990, p. 248) "freedom is actual, or better, actualized, in the interest that we take in the moral law. (...)...the claim is not simply that freedom must be presupposed as a condition of the conceivability of such an interest...(although this is certainly part of the story), it is also that our freedom.. is exhibited in our taking such an interest. (...)...as revealed or manifested in an interest, freedom, like the moral law, is actual only from the practical point of view, since it is only from that point of view that one has interests". Compare Korsgaard 1996a, p. 176: "the point is not that you must *believe* that you are free, but that must choose *as if* you were free".

⁶⁷⁴ As I argued in chapter II, a dogmatic naturalist (qua transcendental realist) can conceive of herself as free (in a weak sense that requires no ontological commitments) and can take a practical interest in morality.

freedom of our theoretical faculties to do some justificatory work for the belief in the freedom of our practical faculties (Wille and Willkür). (1) The belief that we have theoretical freedom must be grounded independently from, and more securely than, the belief in practical freedom. (2) The grounds that support the belief in theoretical freedom must provide grounds for supporting the belief in practical freedom. If condition (1) were not satisfied, the appeal to theoretical freedom would be completely on a par with the appeal to practical freedom and hence could not support it. If condition (2) were not satisfied, the grounds that support the belief in theoretical freedom would be inapplicable to the belief in practical freedom.

There is strong evidence for the idea that in the First Critique, Kant found himself attracted to a strategy of legitimizing the belief in free will that respects these two conditions:

Man is one of the appearances of the sensible world, and in so far one of the natural causes the causality of which must stand under empirical laws. Like all other things in nature, he must have an empirical character. This character we come to know through the powers and faculties which he reveals in his actions. In lifeless, or merely animal, nature we find no ground for thinking that any faculty is conditioned otherwise than in a merely sensible manner. Man, however, who knows all the rest of nature solely through the senses, knows himself also through pure apperception; and this, indeed, in acts and inner determinations which he cannot regard as impressions of the senses. He is thus to himself, on the one hand phenomenon, and on the other hand, in respect of certain faculties the action of which cannot be ascribed to the receptivity of sensibility, a purely intelligible object. We entitle these faculties understanding and reason. The latter, in particular, we distinguish in a quite peculiar and especial way from all empirically conditioned powers. For it views its objects exclusively in the light of ideas, and in accordance with them determines the understanding, which then proceeds to make an empirical use of its own similarly pure concepts. That our reason [diese Vernunft] has causality, or that we at least represent it to ourselves as having causality, is evident from the imperatives which in all matters of conduct we impose as rules upon our active powers. (CPR, A 546-7/B 574-5)

Kant here explicitly states that we have reasons for attributing non-natural capacities only to *some* objects we encounter in nature, namely, to humans rather than animals or lifeless matter. His point cannot be that with regards to animals or inanimate objects we have no grounds for assuming that they have a constitution that differs from their constitution qua empirical object; his idealism requires that *anything* that appears to us is also a thing in itself, that is, has certain properties that differ from the spatiotemporal properties imposed by our cognitive faculties. Rather, Kant's point here is that it is only with regards to human beings that we have reason to think of their non-empirical character in a specific manner, namely, in terms of purely spontaneous powers whose causality does not stand under empirical laws. This confirms that the mere appeal to transcendental idealism (the first dimension of justification) *only makes room for* the belief that we have non-natural causal powers; it does not by itself *provide grounds for adopting* this belief. Now, in this passage Kant specifies that the source of the reasons for attributing to ourselves purely spontaneous faculties is our awareness of the special character of the representations produced by those faculties: namely, their 'purity'. He leaves no doubt that the relevant representations are *theoretical*: they include the pure concepts of the understanding and the ideas of theoretical reason that have an influence or effect on ("determine") the understanding.⁶⁷⁵ Kant moves from the notion that theoretical reason is purely spontaneous to the notion that this same reason ("diese Vernunft") is purely spontaneous also in its practical

⁶⁷⁵ I have tried to spell out what this influence of theoretical reason on the use of the understanding amounts to in chapter XI, sections 3 and 4.

employment ("in allem Praktischen"). Moreover, his claim that we have practical spontaneity has a weaker epistemic status than his claim that we have theoretical spontaneity: Kant modifies the former claim with the qualification that "we at least represent" our practical reason as purely spontaneous, whereas he holds that man *knows himself* through the pure apperception that accompanies the spontaneous mental activities of theoretical reason and the understanding.

A similar line of thought can be witnessed in the Groundwork (GMS, 4: 451-2) where Kant seems to move from the awareness of the purely spontaneous activity of theoretical reason and the understanding to the notion that we are entitled to accept that our will has purely spontaneous causality, and where he seems to suggest that this appeal to the freedom of our theoretical faculties removes the suspicion about a circle in his overall argument.⁶⁷⁶

Before spelling out what I take to be Kant's line of argument here, I want to go back to an important point that came up before (section 2 above; I first discussed this in chapter III, section 3). There is an ambiguity in Kant's conception of practical spontaneity: namely, the distinction between *will* (*practical reason*), as a capacity to produce moral laws, and *Willkür* as a capacity to comply with or to violate those laws.⁶⁷⁷ This is connected to a distinction between spontaneous causality/free agency in a *wide* and in a *narrow* sense. The wide or generic sense relates to any case where an effect is produced by an empirically unconditioned capacity; this includes the production of pure representations, such as moral laws, by pure practical reason. The narrow sense relates to the causal efficacy that *Willkür* exhibits in the choice of maxims and in the production of observable ('outer') effects. The important point, for the purposes of our present discussion is that the wider (generic) sense of spontaneous causality, which applies to the exercise of reason's capacity for legislating practical laws, is not insubstantial: the causal implication of the notion that our practical reason *produces* representations are impossible to deflate, and with regards to this production it is perfectly intelligible to ask (as Kant himself does⁶⁷⁸) whether reason produces its representations (ideas, judgements) from its *self-activity* or under the control of some contingent empirical ground. This point is important for the following reason. The narrow sense of spontaneous causality that relates to the production of observable effects by *Willkür* has no analogue in the case of our theoretical faculties, because unlike the will, our theoretical faculties cannot be considered the cause of the existence of an object (CPR, A 92-3/B 124-5). But the wide or generic sense of purely spontaneous causality or activity applies equally to practical reason, theoretical reason, and the understanding.⁶⁷⁹ Kant claims (cf. the Dialectic passage quoted above) that the understanding qua source of pure concepts is an 'empirically unconditioned power' whose 'causality does not stand under empirical laws'; he affirms (in the parallel Groundwork passage) that the understanding *produces* its pure concepts from its *self-activity*; and he deems both the understanding and practical reason *autonomous*.⁶⁸⁰

These considerations show that the task of legitimizing our belief in free will by appeal to our belief in the freedom of our theoretical faculties is even more complicated than originally

⁶⁷⁶ I discuss the parallelisms between the passages from the Groundwork and the Dialectic in chapter VIII, section 3.

⁶⁷⁷ MS, 6: 226: "Laws arise from the will, viewed generally as practical reason; maxims spring from Willkür."

⁶⁷⁸ Kant says (CPR, A 548/B 576) that natural grounds cannot *produce* a necessary ought judgment and that practical reason literally *makes*, *with spontaneity*, an 'order of ideas' by declaring certain actions necessary.

⁶⁷⁹ There is an important difference here: the categories, unlike the ideas of reason, are confined in their application to what can be given through the senses. But both the ideas of reason and the categories have their origin in pure, sensibly unconditioned spontaneity. See chapter VIII, section 3, and chapter IX, section 6.

⁶⁸⁰ See EEUK, 20: 225; and see my discussion in chapter IX, section 6, and chapter X, section 7.

envisaged. If Kant is to legitimize our belief in free will on the basis of an appeal to theoretical freedom, he must show (1) that the idea that we have the freedom to produce pure *theoretical* representations (concepts, judgments) is more securely grounded than the idea that we have the freedom to produce pure *practical* representations (concepts, judgments); (2) that the appeal to our freedom to produce pure theoretical representations can provide genuine support for the claim that we are free to produce pure practical representations; and (3) that on the basis of (1) and (2) we can support the claim that we also have freedom of Willkür, namely, a capacity to produce effects (i.e., to 'act' in a narrow sense that relates to choice and above all to observable, bodily activity) on the basis of our awareness of pure practical representations (moral laws).

4: The status of the belief in theoretical freedom (freedom of thought)

Let us begin with (1), namely, the credentials for claiming that our epistemic faculties, such as the understanding, are 'empirically unconditioned powers'. Kant holds that this claim *can be inferred* from the special character of the products of our epistemic faculties, i.e., of the conceptual representations that their mental activity produces. As we saw, Kant often refers to this special character in terms of purity (or a priority). What is so special or distinctive about a pure theoretical concept? I suggest that for Kant, a *pure* (a priori) theoretical concept is one that can ground a *necessary* demand that every subject *ought to judge* in a certain manner. Moreover, Kant thinks that we can infer *from* the fact that a cognitive representation (principle) is pure in this sense *to* the fact that the representation is produced in a pure manner, where this means that the production of the representation is not determined by contingent conditions that derive from our empirically given psychology. The idea that one rationally *must* judge in a certain manner is incompatible with supposing that the judgment is founded on concepts or principles that have their origin in our empirically given psychology. Consider, once more, the following passages:

If, however, a judgment gives itself out to be universally valid and therefore asserts a claim to necessity, then, whether this professed necessity rests on concepts of the object a priori or on subjective conditions for concepts, which ground them a priori, it would be absurd...to justify it by explaining the origin of the judgment psychologically. **For...if the attempted explanation were completely successful it would prove that the judgment could make absolutely no claim to necessity, precisely because its empirical origin can be demonstrated.** Now...judgments of reflection...are of the kind mentioned above. They lay claim to necessity and say...that everyone ought so to judge, which is as much as to say that they have an a priori principle for themselves. (EEUK, 20: 239; compare KU, 5: 182)

The universality and necessity in the use of the pure concepts of the understanding betray their origin and that it must be either entirely unreliable and false or else non-empirical. (R 5636, 18: 267-8)

As I argued previously⁶⁸¹, Kant here does *not merely* make the relatively weak point (familiar from Frege and Husserl) that we cannot justify judgments that express necessity by appeal to their psychological origin and causes. Kant's anti-psychologism is much stronger: as the sentence I emphasized (in the first passage) shows, he thinks that the weaker point follows from the stronger point that the empirical origin of a judgment – or of the conceptual representations it employs – proves that the judgment's claim to necessity is null and void.

Kant's point here is completely general: it applies both to the regulative principles of reflective judgment (whose necessity rests on "subjective conditions for [empirical] concepts") and to the constitutive principles of the understanding (whose necessity rests on "concepts of the

⁶⁸¹ See chapter V, section1; and chapter IX, footnote 479.

object a priori"). In the first case, the 'a priori principle' that informs the judgment is exclusively subject-directed: it says that we ought to judge nature (and form our empirical concepts) by assuming, among other things, that nature is a unified system in which the various specific forms of objects (substances, powers) are related in terms of simplicity and unity, even though these assumptions are not constitutive of nature and we do not know that nature will answer to our reflective assumptions. In the second case, the 'a priori principle' of the judgment has both a subject-directed and an object-directed aspect: it informs us both that (for instance) we ought to posit a necessitating causal ground (governed by a general rule) for every natural alteration and that every natural alteration is, as such, caused by a necessitating causal ground (governed by a general rule).⁶⁸² In either case, these 'a priori principles' have necessary validity, and this is incompatible with the supposition that they are produced by empirical, psychological causes. This point underlies Kant's frequent claim that our cognitive commitment to the notion of causal necessity depends for its validity on the idea that it cannot be traced to the contingent, 'unreliable' empirical constitution of the mind (e.g., to the influence of Humean principles of imagination).⁶⁸³

Thus, Kant claims that the pure spontaneity of our epistemic faculties (i.e., their capacity to produce representations/principles without being determined by our empirically given psychological constitution) is a necessary condition of the validity of epistemic principles that are fundamental to our cognitive orientation. We can infer from our awareness of the theoretical necessity expressed by those principles to the pure spontaneity of our cognitive faculties.

I have discussed this aspect of Kant's view at great length elsewhere (cf. chapter IX, section 6; chapter X, sections 5 and 7; chapter XII, sections 3-4). Here I shall only briefly recapitulate one relevant point, which comes out most clearly if we focus on the pure concept of causality and on the corresponding principle that all natural events are subject to law-like necessity. We cannot validate this representation by pointing to anything that the senses present us with (this is Hume's famous point, which Kant wholeheartedly accepts). It is not the 'print' that outer objects impose on our sense organs that justifies the thought that natural events, not to mention *all* such events, are governed by law-like necessity; the *external* world does not force the representation of such necessity on us. Hence, if our thinking in causal terms is exclusively determined by empirical conditions, these conditions must come from the inside, namely, from the empirical constitution of the mind or brain. If our formation and application of the representation of law-like necessity were controlled by the empirical constitution of the mind, we could not regard this representation as having an a priori origin, for the empirical constitution of the mind is a matter of contingent natural law. This dependence on contingent natural causes nullifies the claim to necessity inherent in the principle that every event has a cause. Moreover, if the types of processes that lead us to form and apply the representation of law-like necessity are contingent psychological (or neurological) processes, they cannot be regarded as rational, for the occurrence of these processes is the upshot of (e.g., evolutionary) developments that are (or were) not sensitive to what there is reason to think. Hence, the supposition that our thinking in causal terms is controlled by the empirical constitution of the mind suggests that it stands under the influence of what Hume calls 'custom' or 'habit', and this invalidates our conviction that we are objectively justified in thinking of events as connected through law-like necessity.

⁶⁸² For a detailed discussion of the contrast between these two kinds of principles, see my chapter XI.

⁶⁸³ This is Kant's rejection of Hume's idea that our thinking in causal terms might have its origin in what he calls 'subjective necessity'. See CPR, B 5, B 167-8, A 94/B 126-8. See also footnote 691 below.

It is along these lines that I envisage Kant's defense of the idea that a veridical awareness of the a priori theoretical necessity of cognitive representations (such as the causal principle) presupposes, as its necessary condition, that our higher cognitive faculties produce these representations through purely spontaneous activity that is not a mere function of our contingent psychology (or neurological) constitution. Now, since one philosopher's modus ponens is another's modus tollens, it might be responded that if it is indeed true that the validity of the representation of law-like necessity implicates the pure spontaneity of the faculties that are responsible for this concept, we should reject the notion that this representation *is* valid. This response, however, suggests that our entitlement to the notion that our mind is a part and parcel of nature, and hence wholly conditioned in its activity by empirical laws, is superior to our entitlement to the validity of the representation of law-like necessity. But the notion that our mind is conditioned in its activity by empirical laws *presupposes* the legitimacy of the representation of law-like necessity. So the crux in Kant's defense of the pure spontaneity of the understanding is that one cannot coherently deny the possibility of such spontaneity on the basis of the naturalistic conviction that all our faculties and powers are empirically conditioned by natural laws and causes, because this conviction itself depends on and presupposes the pure spontaneity of the understanding, as a condition of the objective validity of the concept of law-like necessity.⁶⁸⁴ Hence, to argue on empiricist-naturalist grounds that our cognitive faculties stand exclusively under empirical conditions, and (hence) that we have no pure/a priori concepts, would be (to adapt one of Kant's phrases) "as if one sought to prove by reason that there is no reason" (KpV, 5: 12). The naturalistic picture pulls the rug out from under its own feet.

I now want to state how this line of thought combines with the two dimensions of justification that I expounded at the beginning of this chapter. *First*, the notion that the only kind of reality is empirical reality, and that there is no metaphysical space for (empirically) unconditioned causes, is ruled out by transcendental idealism, which makes room for the metaphysical possibility of a capacity for purely spontaneous thought. *Second*, while it must be granted that the idea of epistemic spontaneity violates the conditions of experience, this does not render this idea illegitimate: the spontaneous mind is not an object of empirical knowledge and of scientific theorizing aimed at explaining, observing, and predicting mental activity⁶⁸⁵, but there is a legitimate standpoint other than that of empirical knowledge from which the idea of epistemic spontaneity can be fruitfully employed, namely, the normative standpoint from which one considers how we *ought to* think about nature.⁶⁸⁶ *Third*, what sets the belief that we are capable of spontaneous epistemic activity apart from the belief in the power of foresight is that we are aware of a product that attests to the existence of our capacity for spontaneous activity: we are aware of representations that are necessary conditions for object-directed empirical thought and cognition and that entitle us to judgments about empirical laws and causes in the

⁶⁸⁴ Something analogous might be said about the ideas of theoretical reason (such as systematic unity), which Kant considers to be conditions of systematic scientific reasoning. I have tried to defend the relevance of Kant's point in the context of contemporary discussions of causality or unity in chapter XI, sections 3 and 4.

⁶⁸⁵ This is the point of Kant's claim that "the subject of the categories cannot by thinking the categories acquire a concept of itself as an object of the [schematized] categories" (CPR, B 422; see my chapter X, section 7).

⁶⁸⁶ For a defense of my claim that Kant's non-empirical, normative standpoint addresses the question of how we ought to use both our higher *practical* and our higher *epistemic* faculties, see chapter VIII, sections 3 and 4. See chapter X, sections 4 and 6 for a discussion of how this point applies to the special case of the principles of the understanding, considered both as subject-directed normative conditions for thought and as object-directed metaphysical conditions.

first place. Kant thinks that we *know that* we are capable of purely spontaneous epistemic activity⁶⁸⁷, because such spontaneity is, qua necessary condition of the validity of the representation of (among other things) law-like necessity, a necessary condition of our scientific, causal understanding of nature.

5: The status of the belief in freedom of practical reason and judgment

If the claim that we have a freedom to produce theoretical representations (such as the causal principle) independently of the influence of contingent empirical conditions is defensible, how could this provide support for the idea that we also have a freedom to produce practical representations (such as the categorical imperative) independently of such conditions?

Consider the following passage, the heart of Kant's appeal to the 'fact of reason':

It is therefore the moral law, of which we become directly conscious (as soon as we trace for ourselves maxims of the will), that first presents itself to us, and leads directly to the concept of freedom, inasmuch as reason presents it as a principle of determination not to be outweighed by any sensible conditions, nay, wholly independent of them. But how is the consciousness of that moral law possible? We can become conscious of pure practical laws just as we are conscious of pure theoretical principles, by attending to the necessity with which reason prescribes them and to the elimination of all empirical conditions, which it directs. The concept of a pure will arises out of the former, as that of a pure understanding arises out of the latter. (KpV, 5: 30)

In this passage Kant raises the question of *how consciousness of the moral law is possible*. His answer supplies a criterion for determining when the object of our consciousness is indeed a moral law. So his question is about how we can be certain that we really *are* conscious of a moral law. The previously considered and rejected interpretation that the 'fact of reason' is meant to be entirely self-legitimizing implies that the Kant answers this question by drawing *only* on facets of our moral consciousness. I agree that Kant *does* appeal to such a facet: our awareness of practical necessity. If he left it at that, the important questions about what entitles him to consider the consciousness of such necessity veridical (questions that, as we saw, Kant himself raises in the Groundwork) would remain unanswered. But Kant finds it important to emphasize that the awareness of practical necessity and its connection to our concept of the pure will is analogous to our awareness of theoretical necessity and its connection to our concept of the pure understanding. I want to argue that this analogy provides Kant with a non-dogmatic defense of the veridicality of our awareness of practical necessity.

As we saw in the preceding section, Kant thinks that there is a legitimate inference from the existence of principles that express an a priori theoretical necessity to the pure spontaneity of the epistemic faculty that produces these principles. It is this inference that Kant has in mind in the above passage. The premise of this inference is that there are principles whose claim to necessity is incompatible with the supposition that these principles result from our empirically conditioned mental faculties, that is, from our mind considered as an empirical phenomenon whose activity stands under natural laws. As we also saw, this line of thought relies crucially on the point that the naturalist cannot respond that if this incompatibility is to be conceded, we should give up on the claim that representations such as 'causal necessity' are valid

⁶⁸⁷ See CPR, A 546-7/B 574-5 (as quoted above) for the claim that man "knows himself" as purely spontaneous through pure apperception. See my chapter X, section 7 for a qualified defense of the idea that this claim is compatible with Kant's commitment to 'noumenal ignorance'.

and stick with the claim that our faculty of thought stands exclusively under empirical laws; this response would pull the rug out from under its own feet, because it makes use of the concept of natural causality, which is one of the pure concepts whose legitimacy is at issue here.

Now, we could simply extend the *exact same* line of thought to the pure concepts and principles of practical reason (or the pure will) *if* those concepts and principles were, just as the pure concepts and principles of the understanding, conditions for the possibility of empirical cognition including causal, scientific reasoning. For, under this condition, we could hold that anyone who engages in scientific, empirical reasoning must presuppose as valid the pure concepts and laws of practical reasoning and hence the idea that the exercise of practical reason is not a mere function from empirically given psychological input (i.e., she must presuppose that our practical reason operates autonomously rather than in accordance with a Humean 'desire in, desire out' principle).

There are commentators who endorse this line of thought⁶⁸⁸, but I do not think that such a *direct* defense of pure practical spontaneity by appeal to the conditions of empirical cognition is possible. This is because I do not see any route for arguing that one's causal understanding of the world presupposes that one complies (not only with the causal principle qua necessary condition of object-directed empirical thought but also) with the categorical imperative. One might hold that the quest for empirical cognition requires a commitment to *truth*, and that this gets the subject of empirical cognition into moral territory. But this route would only be open to Kant if he were an epistemic voluntarist, i.e., if he thought that the will can consciously struggle against evidential considerations and influence doxastic states via its conception of what is desirable. That is, if our cognitive commitment to empirical truth were dependent on our commitment to a moral imperative, then, since we (as finitely rational beings) can characteristically violate moral imperatives in full conscious awareness of their validity (cf. chapter VI), we could believe at will contrary to our conscious awareness of the evidential considerations that tell in favor of the empirical truth of some judgment. And this cannot be the default scenario for an anti-voluntarist about belief such as Kant. We cannot choose to be or not to be moved by evidential considerations to believe things about the world (cf. chapter IX).

But the appeal to the conditions of empirical cognition can play a less direct role in the defense of the validity of moral consciousness. Critics of Kant's strategy of trying to support the belief in practical freedom by appeal to theoretical freedom, such as Dieter Henrich, suppose (at least implicitly) that Kant relies on *one* (as they point out, fallacious) inference from theoretical to practical freedom. But I think that at least in the critical writings, such as (most explicitly) the above passage from the Second Critique, Kant is aware that this strategy requires *two separate* inferences: he seeks to infer (first) from the consciousness of the a priori necessity of cognitive principles to the pure spontaneity of our cognitive faculties, and (second) from the consciousness of the a priori necessity of practical principles to the pure spontaneity of the will. I suggest that he intends the former inference to support and shield the latter.

I want to notice right away that I do not conceive of this defense in terms of a strict

⁶⁸⁸ O'Neill 1989, p. 64 argues that the Second Analogy relies on freedom of will. Her reasons for this view are hard to pin down, but a key point seems to be the idea that "agency is here taken to be the presupposition of causal judgment". I agree that for Kant, causal judgment presupposes a kind of agency (see my chapters IX and X), but I do not think that the *epistemic* agency exhibited by the subject of empirical thought presupposes the categorical imperative. Consider: an assassin's non-compliance with moral laws does not count against her causal understanding (achieved through epistemic agency) of the relation between poisoning and killing a victim.

(logical or deductive or transcendental) proof. The soundness of the inference from our awareness of theoretical necessity to the veridicality of this awareness (and hence to the pure spontaneity of the understanding) is compatible with the failure of the inference from the awareness of practical necessity to the veridicality of this awareness (and hence to the pure spontaneity of practical reason). That is: it is possible that while our consciousness of theoretical necessity is veridical, our consciousness of practical necessity is illusory. While (as I argued) there is something self-defeating about the naturalistic attempt to conceive of the thinking mind as governed exclusively by natural causality when the very concept of natural causality seems to presuppose the mind's freedom from nature, there is nothing self-defeating in the imagination of a scenario in which our theoretical faculties are whereas our practical faculties are not capable of purely spontaneous activity. Freedom of will or practical reason can coherently be denied.

But not every philosophical justification needs to be a strict proof. I think that part of what Kant intends with his appeal to the pure spontaneity of our theoretical faculties is a *neutralization* of serious reasons to doubt that we can take our awareness of categorical imperatives at face value. Consider: why would one doubt that our practical reasoning is not a mere function from contingent, empirically given psychological input, and that we can legislate practical laws that express a strict necessity? Certainly, *the major* source of resistance to this idea stems from the naturalistic conviction that we are not capable of producing *any* concepts or principles through a purely spontaneous exercise of our capacities for thought and reasoning because *all* our capacities are empirically conditioned; and this general commitment to the notion that none of our capacities are independent of empirical causes is refuted by the appeal to the fact that our higher *epistemic* faculties *are* capable of a purely spontaneous exercise that results in pure representations. Moreover, to a committed naturalist who holds that *all* our capacities stand under empirical laws, it can be responded that this conception is incoherent, for the very idea that we are wholly and solely subject to natural causes presupposes the legitimacy of the representation of natural, law-like necessity, and this representation cannot be regarded as legitimate if it is attributed to the contingent workings of our empirical mind.

Similarly, why would one claim that we we cannot take the phenomenology of practical necessity that suggests an absence of empirical influence on practical judgment at face value? Certainly, *the major* source of skepticism here is the empiricist conviction that we have no reason to believe in a capacity that cannot be an object of sensible intuition: since we cannot verify the existence of pure practical reason empirically, we can never rule out that the putative effect of the exercise of this capacity, i.e., a judgment that purports to prescribe with strict necessity, is not conditioned by some 'concealed' empirical cause (such as a contingent desire). But again, the appeal to the phenomenology of theoretical necessity shows that we *have* reason to believe in the existence of capacities that cannot be objects of experience, and it preempts the worry that the phenomenology of our awareness of a priori necessity, and the corresponding sense of absence of the influence of contingent psychological causes on judgment, is *generally* untrustworthy.

The line of thought I am attributing here to Kant *turns the tables* on someone asking, 'what reason do I have to take my awareness of the necessity of the moral law at face value, and hence to assume that I have a pure will?' This question can be countered by asking, 'what reason do you have to *doubt* that our practical faculties, like our epistemic faculties, are capable of producing pure principles, when your practical self-conception is inescapably shaped by a consciousness of practical necessity?', and by exposing the standard grounds of suspicion about

the possibility of a pure will and practical necessity as unfounded by an appeal to the actuality of our pure understanding as a source of theoretical necessity. As I said, this does not put the veridicality of moral consciousness beyond *any conceivable* doubt. But it imposes a challenge on those who voice such doubts to *motivate* them in ways which do not fall back on the empiricist and naturalist considerations that tell most strongly against the possibility of practical necessity.

Now, in addition to removing the sting from the naturalistic or empiricist attack on the notion of pure practical reason, the appeal to theoretical necessity also gives us a *positive* reason to treat the consciousness of a priori necessity as veridical. We know that our awareness of a priori necessary laws (such as the causal principle) that inescapably shape our cognitive orientation is veridical: hence, we have a *paradigm case* in which the consciousness of a priori necessity is unassailably certain. Now, in analogy with this paradigm case, we are aware of laws that present themselves to us as having a priori validity and that inescapably shape our practical orientation. It seems in perfect accordance with generally sound principles of reasoning *by analogy* to hold that the paradigm case of awareness of a priori theoretical necessity provides us with positive grounds for trusting our analogous practical self-awareness. The relevant analogy consists in the fact that both the pure understanding and pure practical reason are conceived as autonomous sources of a priori necessities, i.e., as faculties that legislate laws whose validity is essential to our self-conception (as theoretical and practical agents, respectively) without borrowing anything from our contingent internal psychology or from external objects.

To repeat: the reasoning from analogy I am attributing to Kant is not intended to ground demonstrative proofs or deductively valid inferences. Here I think that it is legitimate, rather than evasive, to point out that the standards for the belief in the veridicality of our moral consciousness and hence in the existence of pure practical reason must be attuned to the fact that this consciousness impinges on us from a standpoint of *practical* reasoning. This is not intended to preempt questions about the evidential standards for a belief in freedom altogether, but it legitimizes our dismissal of an insistence that this belief must be supported by the same evidential standards or forms of inference that we rely on in metaphysics or mathematics. The conjunction of the fact that our practical self-conception is strikingly similar in crucial respects to our epistemic self-conception and the fact that the latter self-conception is unassailably veridical provides a reason for taking our practical self-conception at face value that is robust enough *at least* for the purposes of practice.

Thus, the aforementioned conditions (1) and (2) are satisfied: the appeal to the pure spontaneity of our epistemic faculties supports the notion that we can attribute pure 'legislative' freedom to our practical faculties (to produce practical laws) as well. It does so by granting metaphysical and epistemological invulnerability to our consciousness of practical necessity, and by showing that our higher cognitive faculties provide us with an undeniable paradigm case of the veridicality of the awareness of a priori necessity. This allows us to treat the moral law undogmatically as a fact of reason.

6: The status of the belief in freedom of Willkür

Let us now turn to the question of whether the considerations adduced in the two preceding sections can also yield a defense of the idea that we have freedom of *Willkür*. Two features are required for there being a faculty of free Willkür, as Kant conceives of it: for someone who possesses this faculty, it must always be possible (a) to act solely on the basis of her recognition

of the validity of the laws prescribed by practical reason and (b) to act contrary to what is being prescribed by reason, in full awareness that she ought to act in accordance with what practical reason prescribes. (See my chapter VI for an elaboration and defense of this conception.)

If we focus on (a), we must notice (cf. section 2 above) that Kant himself, at least in the *Groundwork* and the *Critique of Practical Reason*, does not worry individually about the possibility of necessary prescriptions of practical reasons (or of pure legislative practical reason) and the possibility that our recognition of the reasons represented by those prescriptions can suffice to determine our actions. He does not entertain the possibility that while practical reason may issue necessary prescriptions, we cannot comply with these prescriptions without drawing in our motivation on a contingent desire. Indeed, it is only in the *Metaphysics of Morals* that he *explicitly* draws the distinction between practical reason and Willkür. So one might complain that Kant did not take seriously enough the alternative that while our practical reason is free to issue necessary prescriptions, the reasons represented by those prescriptions cannot move us to act.

In defense of Kant's dismissal of this alternative, I want to adduce three distinct considerations. The first consideration can be brought out in response to an objection: namely, the worry that the aforementioned distinction between a 'generic' and a 'narrow' sense of spontaneous causality shows that we cannot simply move from the claim that we have pure practical reason to (a). In other words, the worry is that the idea that pure practical reason may all by itself suffice to produce effects in the field of appearances (the narrow conception of spontaneous causality) goes significantly beyond the idea that pure practical reason may spontaneously produce necessary laws (the generic conception of spontaneous causality). But I think that this worry exaggerates the gap between the causality of practical reason and the causality of Willkür. In granting that the generic conception of spontaneous causality is instantiated, one grants that we are capable of an activity (of practical reasoning, that leads to moral judgments) that is not governed by empirical causality. This means that all the general metaphysical resources required for (a) are already in play. Notice here that the causality of Willkür envisaged in (a) relates to two rather different types of effects: first, to the 'inner' act of forming a general maxim or making a concrete choice (e.g., to return money that someone lost); and second, to the 'outer' act that is the intentional object of these inner, intentional states (e.g., returning the money). If one grants that we can reason and judge without depending on the input of contingent desires, it seems a bit arbitrary to insist that when it comes to making choices we can no longer posit such independence from our contingent psychology. And while the capacity to translate our morally good choices into morally good outer actions may in certain cases be impoverished⁶⁸⁹, there is no reason to think that there is a principled problem here.

My second point is that the kind of defense of the freedom of practical reason by appeal to freedom of theoretical thought that I expounded in the preceding section can be extended to the case of Willkür. That is: a principled doubt about the possibility that necessary practical prescriptions may influence actions may be countered by an appeal to the fact that we, in our cognitive practice, presuppose that necessary prescriptions of our epistemic faculties can influence our cognitive acts, namely, our judgments about nature, and that these judgments are not simply the product of contingent empirical processes (such as, say, associative psychological laws). I think that Kant intends to make this point in his review of the essay of Johann Heinrich

⁶⁸⁹ Consider Kant's example of a morally good will that, "owing to...the niggardly provision of a step-motherly nature...should wholly lack power to accomplish its purpose" (GMS, 4: 394).

Schulz, who claimed that there is no free will. Kant says:

...even though he [Schulz] did not concede this to himself, he presupposed that the understanding has the capacity to determine its judgment according to objective reasons that are valid at any time, and does not stand under the mechanism of merely subjectively determining causes, which may change subsequently; he assumed freedom of thought, without which there would be no reason. In the same way [Ebenso] he must presuppose freedom of will for acting, without which there would be no morals [Sitten], if he conducts his...righteous way of life in compliance with eternal laws of duty rather than being a play of instincts and inclinations. (RezSchulz, 8: 14)

Kant does not argue here that the freedom to conduct one's life on the basis of necessary moral reasons *follows from* the concession that we have freedom of thought. Rather, his appeal to freedom of thought has, again, a negative and a positive upshot. The negative point is that our reliance on the idea that our choices can be determined by pure practical reason enjoys a certain kind of invulnerability against the standard naturalistic objections: the naturalist cannot coherently picture us as being *wholly* subject to contingent empirical processes and laws ("the mechanism of merely subjectively determining causes"), for if she did, the very representation of law-like natural necessity that she presupposes in this picture of the human mind and in her particular judgments about the natural causes that determine our mental acts would lose its claim to objective validity. A judgment that a human action is naturally necessitated by some cause is itself the effect of *free, norm-governed epistemic agency*: our commitment to the general notion that there must be a cause for a given action, and our reasoning about what events are to be considered causes of a given action, cannot be attributed to the arbitrary empirical make-up of the mind, or else we lose our grip on the reasons why we should find the deterministic, naturalistic perspective on human action theoretically compelling.⁶⁹⁰ This leaves open the possibility that we lack the analogous practical capacity to choose independently of the influence of our given empirical desires, but we can confront someone who doubts that we have this capacity with the question of what motivates these doubts and with the reminder that these doubts must not, on pain of incoherence, generalize to a full-fledged Humeanism about human reason.⁶⁹¹ The positive point of Kant's line of thought is, again, an inference from analogy. Since we take ourselves to be rationally justified in our judgments about causal relations in nature, we must also take ourselves to be rationally justified in holding that our causal beliefs (and thereby the cognitive commitment to the causal principle that underlies these judgments) are not simply the upshot of contingent empirical processes that might have turned out otherwise and that might change tomorrow. The appeal to our capacity for cognitive self-determination gives us a *paradigm case* of our ability to act (here, judge) solely through our recognition of necessary reasons, independently of the way we happen to be psychologically constituted. Our self-conception as moral agents is inescapably shaped by the analogous conviction that we have reflective control over our choices and outer acts and that we are not, in our choices, the slave of

⁶⁹⁰ See my chapter IX for an elaboration of the notion that judgments about nature are the *effect* of norm-governed epistemic agency. Remember also that Kant does not deny that the naturalistic perspective on human action is valid and compelling if we occupy the empirical standpoint and consider things according to their phenomenal constitution; see my chapter II, section 4 for an explanation of this point.

⁶⁹¹ I find traces of this line of thought in the Preface to the Critique of Practical Reason (KpV, 5: 7-14). Kant notices that the idea of freedom is a "stumbling block" for all empiricists, and he goes on to argue that a thoroughgoing Humean empiricism with respect to human reason (which would seek the foundation of concept of causal necessity in 'subjective necessity', i.e., in contingent psychological habits) is self-refuting.

our contingent desires and passions. So it seems that we can infer, by analogy, that we have good reason to take this self-conception at face value.

I want to adduce a third reason why Kant is justified in assuming that if it is granted that we have pure, legislative practical reason, we can also assume aspect (a) of Willkür. This third reason draws on the general plausibility of some form of internalism about practical reason. The general idea here is that a practical judgment that an agent *can recognize as making a valid demand* on her can also influence her to act. It is difficult, and certainly beyond the scope of this chapter, to provide a general defense of this idea, but it is worth noticing that the most notorious naturalistic skeptics about the possibility that practical reason may motivate the will without being at the service of some contingent desire are *also* skeptics about the existence of categorical imperatives.⁶⁹² This justifies Kant's dismissal of the possibility that we may be capable of pure legislative practical reasoning that yields categorical imperatives but not of (a) determining our choices through our awareness of these imperatives. Hence, if it is possible to preempt naturalistic doubts about the possibility of pure legislative practical reason and about the veridicality of our consciousness of practical necessity, then more than half of the battle about whether we can accept (a), namely, the possibility of desire-independent motivation, might already be won.

I want to back up this idea with a substantive argument. Let us begin by noticing that practical reason can issue valid prescriptions only if it is possible for the addressee to comply with those prescriptions. ('Ought implies can.') Now, if categorical imperatives prescribed merely the production of material effects ('Help the poor'), it would not count against their validity if doing as they say always requires assistance from one or another empirical motive. For in this case all that compliance with the imperative required would be the production of the relevant effect, quite regardless of what motivates the production. But for Kant, the laws prescribed by pure practical reason ask us to adopt general *attitudes* toward people (including ourselves) and to display these attitudes in our outward behavior. The principle that we must always *treat* rational creatures as ends in themselves calls for assigning a certain *unconditional* value to people that is incomparable to the value we may or may not assign to non-rational creatures or material objects by drawing on our contingent empirical preferences. With regards to prescriptions whose content represents unconditional values of this kind, one cannot drive a wedge between one's awareness of these principles as making a valid claim on us and one's awareness of the possibility that these principles can determine us to act independently of sensible motives. Crudely put: one only succeeds in treating rational beings as ends in themselves if one helps those in need out of respect for their humanity rather than – say – because one wants to be perceived as a charitable person. Hence, if we really are aware of valid *oughts* that prescribe that we subordinate our contingent empirical motives to the supreme value of humanity, we are at the same time aware of the fact that we *can* act solely on the basis of the recognition of the supremacy of this value, independently of the influence that contingent motives (such as desires) have on our motivation. This is why, if the veridicality of the consciousness of practical necessity and hence of a pure legislating will can be defended, this also amounts to a defense of one aspect of freedom of Willkür, namely, of (a).

The three considerations I have adduced jointly yield a strong defense of aspect (a) of

⁶⁹² This is certainly true of Hume, but also of Neo-Humeans such as early Foot 1972, Harman 2000, or Williams 1986a, 1997. See my chapter V for further discussion of the Neo-Humean view.

Willkür, namely, of the idea that we can determine our will on the basis of our awareness of necessary practical reasons. How about aspect (b) of Willkür, namely, the idea that we have a pervasive tendency to *violate* the necessary prescriptions of reason? A being that *cannot help but* acting on the basis of the considerations that figure in the content of practical laws is conceivable, but no serious argument is needed to show that this is not the conception we have of ourselves as agents. Kant would treat our awareness that we may fail to be moved by the rational weight of moral norms as a 'datum' that is as basic as the awareness that we may act on the basis of those considerations. But he could also support (b) in a substantive way. Previously (in chapter VI, section 3), I attributed to Kant an argument for (b): Every finite agent is always affected by contingent desires that impel her to 'choose on their behalf', i.e., to make their satisfaction a condition of choice; practical norms give necessary, desire-independent reasons for acting; hence, every finite agent is always tempted in her choices to resist governance by (the reasons provided by) practical norms

7: Summary: Kant's justification of the belief in practical freedom

I now want to summarize how I conceive of Kant's justification of the belief that we have practical freedom, i.e., that we have free will (practical reason and Willkür).

First, there is metaphysical space, opened up by transcendental idealism, for the possibility of the will qua unconditioned causality; the notion that every cause stands under a determining cause in time is an epistemic condition for the possibility of experience, but not an ontological condition of being or acting as such. So, adopting the belief that our will is free at the noumenal level does not contradict our belief that the will (qua appearance) is causally determined. Second, while the belief in freedom of will violates the conditions of the possibility of experience, we adopt it in a context where we are liberated from the conditions of experience and from the standards that govern the adoption of empirical belief because we do not aim at cognizing the properties of empirical objects: we are not interested in how things are, or will be, or why they happen as they do, but in specifying how people ought to behave.

These two points are not sufficient for supporting the belief that we have free will rather than the power of foresight. The reasons we have for adopting this belief derive from the reasons we have for taking at face value our consciousness of practical necessity. *If* we can take this awareness of practical necessity at face value, as a *veridical* consciousness of the objective validity of necessary prescriptions, it follows that we have reason to believe, first, in the existence of pure practical reason and, second, in the existence of free Willkür. The first belief would follow from the veridical awareness of practical necessity because there can only be necessary practical principles if we can reason practically in a way that is not a Humean function from empirically given desires. The belief in a free Willkür comprises (a) the belief in a capacity to comply with necessary practical norms and (b) the belief in a pervasive tendency to violate these norms. The belief in (a) the capacity to comply with such norms would be supported by the fact that we have a veridical awareness of practical necessity for the following reason: we have a veridical awareness of valid *oughts* only if we *can* comply with them; moral principles prescribe acts that display respect for an *unconditional* value; hence we can comply with these prescriptions *only if* we can act on the basis of the recognition of the supremacy of this value, without motivationally depending on some conditional desire. It follows that the veridicality of our awareness of valid moral norms requires that our recognition of the necessary reasons that

these norms represent can all by itself lead us to act. The belief in (b) the tendency to violate necessary practical norms rests on the combination of our veridical awareness of norms that prescribe choosing for desire-independent reasons and the undeniable fact that we are always affected by contingent desires that impel us to choose on their behalf.

We *do* have reason to take our awareness of valid, a priori necessary practical principles at face value. This is so because the belief in the veridicality of this consciousness has a theoretical analogue, the awareness of a priori theoretical necessity. The fact that the theoretical necessities represented by our cognitive faculties and (thereby) the pure spontaneity of those faculties are preconditions of empirical cognition shows that the general naturalist and empiricist attempt to disenchant and vilify the consciousness of a priori practical necessity fails. So the belief in the veridicality of this consciousness is metaphysically and epistemologically invulnerable against its most serious challengers. Moreover, the fact of a priori theoretical necessity provides us with a paradigm case for taking the consciousness of a priori necessity at face value that yields a positive reason for believing in the veridicality of our analogous awareness of a priori practical necessity. This does not demonstrably prove that this awareness is veridical, but the request for a demonstrative proof of the legitimacy of our moral self-conception can be dismissed by appeal to the fact that we are not endorsing this self-conception for those theoretical (e.g., metaphysical) purposes that would call for strict demonstrability.

This reconstruction of Kant's justification of our belief in practical freedom removes the air of dogmatism that seems to surround the appeal to the (consciousness of the) moral law as a fact of reason. I am not, of course, suggesting that the appeal to theoretical necessity accompanies the consciousness of moral necessity, or that the reasons Kant (on my interpretation) supplies for taking the awareness of practical necessity at face value are typically displayed in our moral consciousness. These are reasons that could be cited in defense of our moral consciousness upon reflection; and they explain why our consciousness of the moral law can be considered, in Kant's famous words, the *ratio cognoscendi* of free will:

Lest any one should imagine that he finds an inconsistency here when I call freedom the condition of the moral law, and hereafter maintain in the treatise itself that the moral law is the condition under which we can first become conscious of freedom, I will merely remark that freedom is the ratio essendi of the moral law, while the moral law is the ratio cognoscendi of freedom. For had not the moral law been previously distinctly thought in our reason, **we should never consider ourselves justified in assuming such a thing as freedom**, although it be not contradictory. But were there no freedom it would be impossible to trace the moral law in ourselves at all. (KpV, 5: 4; my emphasis)

This passage shows, I think, that attempts to represent Kant as being concerned only with choice, practical needs and interests rather than *justified assent* to the proposition *that we are free* are misguided. The very notion of *ratio cognoscendi* signals that Kant wants to hold that the moral law is a *ground of cognition* of freedom, a *source of evidence* for the belief in freedom which allows us to "cognize" (erkennen) our freedom (KpV, 5: 30).⁶⁹³ We must keep this point in mind when we turn to Kant's insistence that the moral law allows for a deduction of freedom that gives the concept of freedom *objective reality which is of a (merely) practical nature* (KpV, 5: 47 ff.). He does not mean here that the moral law gives us a practical reason for the belief in freedom.

⁶⁹³ I agree with Hogan 2009 (against Chignell 2007a, 2007b) that this practical evidence affords us with genuine *knowledge* that we are free. As Hogan points out, for Kant the term 'knowledge' is not limited to the knowledge we acquire on the basis of theoretical (e.g., empirical) evidence.

Practical reasons pertain to what it is rational to *choose*; a reason for adopting a doxastic state (what we would call belief, and what Kant calls 'assent') toward the proposition that we have practical freedom must be an *evidential* consideration. An exercise of the will is not itself responsible for the formation of the belief that we have practical freedom; rather, an exercise of the will *qua* practical reason provides us with *evidence* for this belief, since the exercise of pure practical reason affords us with the consciousness of necessary practical laws that serves as a legitimate ground for adopting the belief in free will. We should understand Kant's claim that the concept of freedom has objective, but only *practical* reality as suggesting that the *only* kind of evidence we have for our belief in freedom of will stems from our veridical awareness of a *distinctively practical* necessity, i.e., a consciousness of unconditional reasons for acting that impinges on us only when we face questions about how one ought to choose.⁶⁹⁴

Although Kant speaks of a *deduction* of freedom by appeal to the moral law, he clearly does not envisage here a strict transcendental deduction on par with that of the categories. The very notion of a deduction of freedom on the basis of the moral law is informed by the idea that the moral law gives the concept of freedom only practical objective reality (cf. KpV, 5: 47). We should understand him as claiming that the moral law gives us a kind of evidence for the belief in practical freedom that suffices to license this belief given the evidential standards associated with the standpoint from which the consciousness of the moral law impinges on us. That is: the robust, persistent, and inescapable phenomenology of practical necessity licenses – given the analogous, securely grounded inference from theoretical necessity to epistemic spontaneity – an inference to practical spontaneity as the ontological foundation, *the ground of being* or *ratio essendi*, of the practical necessity represented in the moral law.

The account of the justification of the belief in practical necessity and freedom I envisage has an important upshot: it implies that our self-conception as subjects of practical and theoretical reasoning is, in a deep sense, *unified* insofar as it is *one* generic kind of picture of ourselves that we can, on reflection, endorse as a precondition of the possibility of both theory and practice⁶⁹⁵: namely, the view of ourselves as autonomous agents who impose laws on thought and volition and who have a (fallible) reflective sovereignty to determine their beliefs and choices through their recognition of the validity of the reasons represented by these laws. This provides final support for my claim (that I first put forward in chapter VIII) that for Kant the *standpoint* that contrasts with the empirical perspective on ourselves as naturally necessitated parts of nature is not just a standpoint of practical reasoning but a perspective that considers us as free norm-governed practical *and* epistemic agents, and that aims at establishing how we *ought to choose and think*.

⁶⁹⁴ Cf. KpV, 5: 30: "...no one would ever have been so rash as to introduce freedom into science, had not the moral law, and with it practical reason, come in and forced this notion upon us."

⁶⁹⁵ However, my interpretation should be distinguished from the 'coherentist' interpretations provided by Beck (1960, p. 174) and Rawls (1989, p. 108). On their account, the justification of the moral law hinges on the fact that, in providing reality for the idea of the unconditioned, the moral law answers to a need of theoretical reason, namely, theoretical reason's need for a positive determination of its abstract idea of the unconditioned. Kant indeed holds that this point lends *additional* support to the concept of the moral law (KpV, 5: 47). But I would not want to put this point at the centre of Kant's line of thought in the Second Critique, because the argument for the moral law as a *ratio cognoscendi* of freedom comes much earlier than, and proceeds independently of, the appeal to the fact that the moral law gives reality to an idea that theoretical reason was compelled to assume. Moreover, the idea that the moral law can provide reality for the idea of the unconditioned *already presupposes* that it has been shown that our consciousness of moral law can be accepted as veridical.

I want to conclude by addressing two kinds of worries that one might raise about the interpretation I have put forward in this chapter.

The first type objection is that my frequent appeal to the analogy between our awareness of theoretical and practical laws ignores crucial *disanalogies*, such as the fact that the reaction to the moral evil cannot be equated with the reaction to theoretical error.⁶⁹⁶ My response is that this would be an objection to Kant's 'inference from analogy' only if he were to suppose that the inference is grounded on an analogy between theoretical and practical error. While Kant may have been flirting with this analogy in some of his early lectures, this is not *the* analogy that I find him relying on in the major critical writings and that has informed my argument in this chapter, namely, the analogy that derives from the fact both our practical and our epistemic orientation is permeated with a sense of necessary demands on our agency and that pure spontaneity is a necessary condition for the validity of these demands.

One might follow up with the worry that the sources of theoretical and practical necessity are wholly different: the source of theoretical necessity is that compliance with the relevant laws is a condition of empirical thought or cognition, and nothing analogous can be said in the practical case. I accept this point (i.e., I am deeply skeptical about interpretations that portray Kant as holding that complying with moral laws is in some sense constitutive of willing as such, rather than of good willing⁶⁹⁷). But this point *would* be an objection to Kant's inference from analogy, or to my reconstruction of the argumentative weight of this inference, *only if* the inference were meant to show that we have unconditional reasons to comply with moral norms. Here we must carefully distinguish between two senses of a 'justification' of the moral law: first, the justification of the claim that we can take at face value the awareness of unconditioned practical necessities, in view of the fact that the veridicality of this awareness implies the falsity of a universally naturalistic and empiricist conception of human reason; second, the justification of the claim that we have reasons to comply with moral laws. The first justification is metaphysical and epistemological, the second one is normative. I have been concerned here *exclusively* with the first sense in which (our consciousness of) the moral law, and the conception of practical reason that underlies it, stands in need of justification. Now, I think that at least in the *Critique of Practical Reason*, Kant gives up on any attempt to provide a 'justification of morality' in the *second* sense described above. What is distinctive of the appeal to a fact of reason here is that he takes our acceptance of the validity of the demand the moral law makes on us to be a primitive part of ordinary moral consciousness that cannot be given further justification (at least no justification that could add to the prescriptive normative force that is already inherent in fundamental moral ideas as we become aware of them independently of philosophical reflection).⁶⁹⁸ The philosophical defense Kant gives of this fact pertains not to its normative

⁶⁹⁶ See Henrich 1954-5, pp. 29-34.

⁶⁹⁷ This seems to be a theme in Korsgaard 2009.

⁶⁹⁸ This is the point of Kant's claim (KpV, 5: 47) that the moral law itself needs no justifying a priori reasons. Compare KpV, 5: 32: "The fact just mentioned is undeniable. It is only necessary to analyze the judgement that men pass on the lawfulness of their actions, in order to find that, whatever inclination may say to the contrary, reason, incorruptible and self-constrained, always confronts the maxim of the will in any action with the pure will, that is, with itself, considering itself as a priori practical) and KpV, 5: 35 ("...and would ruin morality altogether were not the voice of reason in reference to the will so clear, so irrepressible, so distinctly audible, even to the commonest men"). Kant attempts to give an account of the motivating reasons and interests that are operative in a pure moral will, but here he does not try to *establish that we ought to* have respect for the moral law but to provide an analysis of our 'primitive' sense of respect for the moral law and a defense of the claim that

credentials but to its metaphysical foundations (its *ratio essendi*).

Now, one can simply recast, *at the normative level*, the charge of dogmatism that I have tried to acquit Kant of at the metaphysical and epistemological level. How can Kant just assume that moral laws make necessary demands without providing a deduction of the values and reasons represented by those demands? It would require a separate paper, or even book, to adequately address this worry. On Kant's behalf, it can be said here that it is not clear that any philosophical defense can really add substantively to the intuitive appeal that, say, the notion that rational beings are ends in themselves has, at least not in a way that would be able to convince an amoralist. Of course the appeal to the possibility of amoralism must be taken seriously as an attack on Kant's notion that awareness of the moral law is a timeless, necessary feature of the self-consciousness of rational practical agency as such. One would have to consider carefully whether the amoralist really does, consistently, reject the categorical imperative as false or rather considers herself as complying with this imperative in some delusional manner. However this may be, it should be obvious that these worries are completely separate from the important epistemological and metaphysical worries against which Kant's appeal to the fact of reason does, as I have argued, provide a non-dogmatic defense.

A second kind of worry about my interpretation relates to my claim that the reasons for believing that we are capable of pure epistemic spontaneity are stronger, or more conclusive, than the reasons for believing that we are capable of pure practical spontaneity. One might worry that given this dialectical situation, it is somewhat odd for Kant to say that the concept of transcendental freedom has 'only' practical reality. Given the firm grounds for supposing that we have absolute freedom of thought, shouldn't he be saying that this concept has first and foremost a kind of 'epistemic' or 'theoretical' reality?

Here a couple of points must be acknowledged. First, as we saw, for Kant the notion of spontaneous causality has both a narrow and a wide sense, and the narrow sense – which relates to the production of observable effects in the field of appearances – pertains only to the spontaneous causality of Willkür. This narrow sense is indispensable as far as our practical self-conception is concerned, because 'outer' effects in the field of appearances are what we intend in practical deliberation, and such effects are our only (inconclusive) basis for inferring to the moral character of agents. Hence, our (justified) conviction that we can realize the objects of our pure moral intentions in experience provides a kind of 'reality' to the notion of freedom or spontaneity that cannot be gained from our awareness as thinkers who stand under necessary laws of theoretical reason. Second, for Kant the concept of transcendental freedom, as applied to the will of imperfectly rational beings, has a very, indeed *regrettably* important component that has no analogue in the case of the freedom of our cognitive faculties: namely, what I referred to as aspect (b) of Willkür, which accounts for our tendency to choose in ways that fail to accord with our conscious awareness of how we ought to choose. It is only via our consciousness of the moral law qua moral *ought*, and hence through the distinctively practical reality provided for the concept of transcendental freedom by this consciousness, that we have reason to posit this tendency. Third, Kant's conception of the reality of the idea of freedom relates crucially to the *interest* that we take in this idea. And whatever we can say about the credentials for the belief in

this respect can be considered a pure, sensibly unconditioned *Triebfeder* of moral agency. That Kant is not trying to demonstrate or 'derive' (i.e., deduce) the absolute priority of moral value is emphasized by Guyer 1995, p. 375 and by Herman 1996, p. 240. Unlike Guyer, I am not quite convinced that Kant wholly renounced this goal already in the Groundwork, but this does not matter for present interpretative purposes.

pure epistemic spontaneity, and however necessary pure epistemic spontaneity may be for empirical cognition, the *interest* we have in our freedom is predominantly of a practical nature: it is in the practical sphere that the question of whether or not we are free or spontaneous *really, existentially*, matters. Moral questions matter to us above all other questions, and our sense that people are morally responsible for what they are doing is (whereas our sense that people are epistemically responsible for what they believe is not) an essential component of coexistence.⁶⁹⁹ Moreover, Kant thinks that the belief in practical freedom, *unlike* the belief in our epistemic spontaneity⁷⁰⁰, may furnish us with a way to defend the claim that our souls are distinct from our bodies, (hence) that we are immortal, and that God (as an all-benevolent being that provides happiness for those, and only for those, who are worthy of it) exists. And I think that he would say that it is our awareness of the urgency and importance of these issues that gives a kind of 'reality' to the concept of practical freedom that makes its epistemic counterpart appear utterly pale by comparison.

⁶⁹⁹ See chapter IX, footnote 480 for the suggestion that our sense of responsibility for doxastic states is thinner than our sense of responsibility for volitional states (since only our concept of moral responsibility concerns ideas of praise and blame that are based on respect or disrespect for humanity or for rational beings).

⁷⁰⁰ See chapter X, section 7.

Bibliography

Quotations from Kant's works, apart from the Critique of Pure Reason (CPR), cite the volume and page number of the Academy edition (Gesammelte Schriften ed. Königlich Preussische Akademie der Wissenschaften (Berlin: Walter de Gruyter, 1902-, Volumes 1-29). Quotations from the Critique of Pure Reason are cited according to the standard A and B pagination for the first and second editions, respectively. The translations I used for quotations are indicated below.

Abela, Paul (2002): *Kant's Empirical Realism*; Oxford University Press

Allais, Lucy (2010): "Transcendental Idealism and Metaphysics: Kant's Commitment to Things as They are In Themselves"; in: Dietmar H. Heidemann (ed.), *Kant Yearbook 2010: Metaphysics*, pp. 1-33. Walter De Gruyter

Allison, Henry (1989): "Kant and the Claims of Knowledge" (Review); in *Journal of Philosophy* 86 (2), pp. 214-22

Allison, Henry (1990): *Kant's Theory of Freedom*; Cambridge University Press

Allison, Henry (1996): *Idealism and Freedom*; Cambridge University Press

Allison, Henry (2004): *Kant's Transcendental Idealism*; Yale University Press

Alston, William (1989): "Level Confusions in Epistemology"; in Alston, *Epistemic Justification: Essays in the Theory of Knowledge*, pp. 153-71. Cornell University Press

Ameriks, Karl (1992): "Kantian Idealism Today"; in *History of Philosophy Quarterly* 9, pp. 329-42

Ameriks, Karl (2000): *Kant's Theory of Mind*; Oxford University Press

Anderson, Lanier R. (2001): "Synthesis, Cognitive Normativity, and the Meaning of Kant's Question, 'How Are Synthetic Cognitions A Priori Possible?'"; in *The European Journal of Philosophy* 9 (3), pp. 275-305

Anscombe, G.E.M. (1958): "Modern Moral Philosophy"; *Philosophy* 33, Nr. 124

Anscombe, G.E.M. (1993): "Causality and Determination"; in M. Tooley and E. Sosa (eds.), *Causation*, pp. 88-104. Oxford University Press

Aquila, Richard (1979): "Things in Themselves and Appearances: Intentionality and Reality in Kant"; in: *Archiv für Geschichte der Philosophie*, 61 (1979), pp. 293-308

Aquila, Richard (1981): "Intentional Objects and Kantian Appearances"; in: *Philosophical Topics*, 12/2 , pp. 9-37

Aristotle (EN): *Nicomachean Ethics*; The Complete Works of Aristotle Volume 2: The Revised Oxford Translation, Jonathan Barnes (ed.); Princeton University Press

Aristotle (EE): *Eudemian Ethics*; The Complete Works of Aristotle Volume 2: The Revised Oxford Translation, Jonathan Barnes (ed.); Princeton University Press

Armstrong, David (1983): *What is a Law of Nature?*; Cambridge University Press

Ayer, A.J. (1963): "What is a law of nature?"; in Ayer, *The Concept of a Person and other Essays*. Macmillan

Beck, Lewis White (1960): *A Commentary on Kant's 'Critique of Practical Reason'*; University of Chicago Press

Beck, Lewis White (1987): "Five Concepts of Freedom in Kant"; in J. T. J. Szrednick (ed.), *Philosophical Analysis and Reconstruction: a Festschrift to Stephan Körner*, pp. 35-51. Martinus Nijhoff

Baumgartner, Michael (2008): "Regularity Theories Reassessed"; in: *Philosophia* 34, pp. 327-54

Beebe, Helen (2003): "Local Miracle Compatibilism"; *Nous* 37: 2, pp. 258-77

Bennett, Jonathan (1966): *Kant's Analytic*; Cambridge University Press

Bennett, Jonathan (1974): *Kant's Dialectic*; Cambridge University Press

Bennett, Jonathan (1984): "Commentary: Kant's Theory of Freedom"; in Allen Wood (ed.), *Self and Nature in Kant's Philosophy*, pp. 102-12. Cornell University Press

Bennett, Jonathan (1990): "Why is Belief Involuntary?"; in *Analysis* 50, pp. 87-107

Bentham, Jeremy (2007): *An Introduction to the Principles of Morals and Legislation*; Dover Publications

Bird, Graham (1962): *Kant's Theory of Knowledge*; Routledge Kegan Paul

Bittner, Rüdiger (1983): *Moralisches Gebot oder Autonomie*; Alber

Bojanowski, Jochen (2006): *Kant's Theorie der Freiheit*; Walter De Gruyter

Bok, Hilary (1998): *Freedom and Responsibility*; Princeton University Press

- Boyd, Richard (1988): "How to Be a Moral Realist"; in G. Sayre McCord (ed.), *Essays on Moral Realism*, pp. 181-228. Cornell University Press
- Broad, C.D. (1978): *Kant: An Introduction*; Cambridge University Press
- Brandt, Reinhard (1988): "Der Zirkel im dritten Abschnitt von Kants Grundlegung zur Metaphysik der Sitten"; in H. Oberer and G. Seel (eds.), *Kant: Analysen-Probleme-Kritik*, pp. 169-91. Königshausen und Neumann
- Brandt, Richard (1979): *A Theory of the Good and the Right*; Oxford University Press
- Bratman, Michael (1987): *Intention, Plans, and Practical Reason*; Harvard University Press
- Brewer, Bill (1995): "Mental Causation: Compulsion By Reason"; *Proceedings of the Aristotelian Society Supplementary Volume 69*, pp. 237-53
- Brook, Andrew (1994): *Kant and the Mind*; Cambridge University Press
- Broome, John (2007): "Wide or narrow scope?"; in *Mind* 116, pp. 359-70
- Broome, John (2009): "The unity of reasoning?"; in Simon Robertson (ed.), *Spheres of Reason*, pp. 62-92. Oxford University Press
- Buchdahl, Gerd (1974): "The Conception of Lawlikeness in Kant's Philosophy of Science"; in L.W. Beck (ed.), *Kant's Theory of Knowledge*, pp. 128-50. Reidel
- Burge, Tyler (1998): "Reason and the First Person"; in C. Wright, B. Smith and C. McDonald (eds.), *Knowing Our Own Minds*, pp. 243-70. Oxford University Press
- Carnois, Bernard (1987): *The Coherence of Kant's Doctrine of Freedom*. University of Chicago Press
- Chignell, Andrew (2007a): "Belief in Kant"; in *Philosophical Review*, Vol. 116, pp. 323-60
- Chignell, Andrew (2007b): "Kant's Concepts of Justification"; in *Nous* 41:1, pp. 33-63
- Chisholm, Roderick M. (2003): "Human Freedom and the Self"; in Gary Watson (ed.), *Free Will*, pp. 26-37. Oxford University Press
- Cramer, Konrad (1972): "Hypothetische Imperative?"; in Manfred Riedel (ed.), *Rehabilitierung der praktischen Philosophie*, Vol. 1, pp. 159-212. Rombach
- Davidson, Donald (1973): "Freedom to Act"; in Ted Honderich (ed.), *Essays on Freedom of Action*, pp. 137-56. Routledge and Kegan Paul

Davidson, Donald (2001): "Mental Events"; in: Davidson, *Essays on Actions and Events*. Oxford University Press

Dennett, Daniel (1984): "I Could Not Have Done Otherwise – So What?"; in *Journal of Philosophy* 81, pp. 553-65

Dickerson, A.B. (2007): *Kant on Representation and Objectivity*; Cambridge University Press

Ducasse, C.J. (1993): "On the Nature and the Observability of the Causal Relation"; in M. Tooley and E. Sosa (eds.), *Causation*, pp. 125-36. Oxford University Press

Dretske, Fred (1977): "Laws of Nature"; in *Philosophy of Science* Vol. 44, No. 2, pp. 248-68

Earman, John (2004): "Determinism: What We Have Learned and What We Still Don't Know"; in Joseph Campbell, Michael O' Rourke and David Shier (eds.), *Freedom and Determinism*, pp. 21-46. MIT Press

Engstrom, Stephen (1994): "The Transcendental Deduction and Skepticism"; in: *Journal of the History of Philosophy* 32, pp. 359-80

Fales, Evan (1990): *Causation and Universals*; Routledge

Falkenstein, Lorne (1995): *Kant's Intuitionism. A Commentary on the Transcendental Aesthetic*; University of Toronto Press

Findlay, J.N. (1981): *Kant and the Transcendental Object*; Clarendon Press

Foot, Philippa (1972): "Morality As A System of Hypothetical Imperatives"; in *The Philosophical Review* Vol. 81 (No. 3), pp. 305-16

Forschner, Maximilian (1974): *Gesetz und Freiheit. Zum Problem der Autonomie bei I. Kant*; Pustet

Frankfurt, Harry (1988a): "The importance of what we care about"; in Frankfurt, *The Importance of What We Care About. Philosophical Essays*, pp. 80-94. Cambridge University Press

Frankfurt, Harry (1988b): "Identification and wholeheartedness"; in Frankfurt, *The Importance of What We Care About. Philosophical Essays*, pp. 159-76. Cambridge University Press

Frankfurt, Harry (2003a): "Freedom of the Will and the Concept of a Person"; in Gary Watson (ed.), *Free Will*, pp. 322-36. Oxford University Press

Frankfurt, Harry (2003b): "Alternate Possibilities and Moral Responsibility"; in Gary Watson (ed.), *Free Will*, pp. 167-76. Oxford University Press

- Freudiger, Jürg (1993): *Kant's Begründung der praktischen Philosophie. Systematische Stellung, Methode und Argumentationsstruktur der 'Grundlegung der Metaphysik der Sitten'*; Haupt
- Friedman, Michael (1992a): "Causal laws and the foundations of natural sciences"; in Paul Guyer (ed.), *The Cambridge Companion to Kant*, pp. 161-99. Cambridge University Press
- Friedman, Michael (1992b): *Kant and the Exact Sciences*; Harvard University Press
- Friedman, Michael (1992c): "Regulative and Constitutive"; in *The Southern Journal of Philosophy* 30 (S1), pp. 73-102
- Friedman, Michael (1996): "Exorcising the Philosophical Tradition: Comments on John McDowell's 'Mind and World'"; in *The Philosophical Review* 105, pp. 427-67
- Friedman, Michael (1999): "Carnap's *Aufbau* Reconsidered"; in: Friedman, *Logical Positivism Reconsidered*, pp. 89-113; Cambridge University Press
- Funke, Gerhard (1981): "Kants Satz: Die praktische Freiheit kann durch Erfahrung bewiesen werden"; in: *The Philosophical Quarterly* 31, pp. 207-21
- Gibbard, Allan (2011): *Reconciling Our Aims: In Search of Bases for Ethics*, Barry Stroud (ed.); Oxford University Press
- Ginsborg, Hannah (1991): "Korsgaard on Choosing Nonmoral ends"; in *Ethics* 109 (1), pp. 5-21
- Ginsborg, Hannah (1998): "Kant on the Subjectivity of Taste"; H. Parret (ed.), *Kants Ästhetik/ Kant's Aesthetics/ L'esthétique de Kant*, pp. 448-65. Walter de Gruyter
- Ginsborg, Hannah (2006): "Kant and the Problem of Experience"; in *Philosophical Topics* (1&2), pp. 59-106
- Ginsborg, Hannah (2008): "Was Kant a Nonconceptualist?"; in *Philosophical Studies* 137 (1), pp. 65-77
- Goldman, Alvin (1967): "A Causal Theory of Knowing"; in *The Journal of Philosophy* 64, pp. 357-72
- Goldman, Alvin (1986): *Epistemology and Cognition*; Harvard University Press
- Goodman, Nelson (1955): *Fact, Fiction and Forecast*; Harvard University Press
- Grier, Michelle (2001): *Kant's Doctrine of Transcendental Illusion*; Cambridge University Press

Gunkel, Andreas (1989): *Spontaneität und moralische Autonomie: Kants Philosophie der Freiheit*; Haupt

Guyer, Paul (1980): "Kant on Apperception and 'A Priori' Synthesis"; in *American Philosophical Quarterly* Vo. 17, No. 3, pp. 205-12

Guyer, Paul (1987): *Kant and the Claims of Knowledge*; Cambridge University Press

Guyer, Paul (1990): "Reason and Reflective Judgment: Kant on the Significance of Systematicity"; in *Nous*, Vol. 24 No. 1, pp. 17-43

Guyer, Paul (1992): "Introduction"; in Guyer (ed.), *The Cambridge Companion to Kant*, pp. 1-26. Cambridge University Press

Guyer, Paul (1995): "The Possibility of the Categorical Imperative"; in *The Philosophical Review* Vol. 104 No. 3, pp. 353-85

Guyer, Paul (2003): "Kant on the Systematicity of Nature: Two Puzzles"; in *History of Philosophy Quarterly*, Vol. 20 No. 3, pp. 277-95

Hacking, Ian (1991): "A Tradition of Natural Kinds"; in *Philosophical Studies* 61, pp. 109-26

Hampshire, Stuart (1971): *Thought And Action*; The Viking Press

Hampton, Jean (1998): *The Authority of Reason*; Cambridge University Press

Hanna, Robert (2005): "Kant and nonconceptual content"; in *European Journal of Philosophy*, 13(2), pp. 247-90

Hanna, Robert (2006): *Kant, Science, and Human Nature*; Clarendon Press.

Harman, Gilbert (1977): *The Nature of Morality: An Introduction to Ethics*; Oxford University Press

Harman, Gilbert (2000): "What Is Moral Relativism?"; in Harman, *Explaining Value and Other Essays in Moral Philosophy*, pp. 20-38. Oxford University Press

Hatfield, Gary (1990): *The Natural and the Normative: Theories of Spatial Perception from Kant to Helmholtz*; MIT Press

Hatfield, Gary (1992): "Empirical, rational and transcendental psychology: Psychology as science and as philosophy"; in Paul Guyer (ed.), *The Cambridge Companion to Kant*, pp. 200-27. Cambridge University Press

- Heal, Jane (1996): "Simulation, Theory and Content"; in P. Carruthers and P. K. Smith (eds.), *Theories of Theories of Mind*, pp. 75-89. Cambridge University Press
- Henrich, Dieter (1954-5): "Das Prinzip der kantischen Ethik"; in *Philosophische Rundschau* 2, pp. 29-34
- Henrich, Dieter (1969): "The Proof-Structure of Kant's Transcendental Deduction"; in *Review of Metaphysics* 22, pp. 640-59
- Henrich, Dieter (1973): "Der Begriff der sittlichen Einsicht und Kants Lehre vom Faktum der Vernunft"; in G. Prauss (ed.), *Kant: Zur Deutung seiner Theorie vom Erkennen und Handeln*, pp. 77-115. Kiepenhauer & Wietsch
- Henrich, Dieter (1975): "Die Deduktion des Sittengesetzes"; in A. Schwan (ed.), *Denken im Schatten des Nihilismus*, pp. 55-112. Wissenschaftliche Buchgesellschaft
- Henrich, Dieter (1976): *Identität und Objektivität. Eine Untersuchung über Kants transzendente Deduktion*; Carl Winter
- Henrich, Dieter (1989): "Kant's Notion of a Deduction and the Methodological Background of the First Critique"; in E. Förster (ed.), *Kant's Transcendental Deduction*, pp. 29-46. Stanford University Press
- Henson, Richard G. (1979): "What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action"; in *The Philosophical Review* 88, pp. 39-54
- Herman, Barbara (1996): *The Practice of Moral Judgment*; Harvard University Press
- Herman, Barbara (2001): "Rethinking Kant's Hedonism"; in R. Stalnaker, R. Wedgwood and A. Byrne (eds), *Fact and Value: Essays on Ethics and Metaphysics for Judith Thomson*, pp. 129-54. MIT Press
- Hieronymi, Pamela (2009): "Two Kinds of Agency"; L. O' Brien and M. Soteriou (eds.), *Mental Actions*, pp. 138-62. Oxford University Press.
- Hieronymi, Pamela (2006): "Controlling Attitudes"; *Pacific Philosophical Quarterly* 87, pp. 45-74
- Hill, Thomas (1973): "The Hypothetical Imperative"; in *The Philosophical Review* 82, pp. 429-50
- Hill, Thomas (1992): *Dignity and Practical Reason in Kant's Moral Theory*; Cornell University Press

Hinske, Norbert (1995): "Prolegomena zu einer Entwicklungsgeschichte des Kantschen Denkens. Erwiderung auf Lothar Kreimendahl"; in R. Theis and Claude Weber (eds.), *De Christian Wolff à Louis Lavelle: métaphysique et histoire de la philosophie; recueil en hommage à Jean Ecole à l'occasion de son 75e anniversaire*. Olms Verlag

Hogan, Desmond (2009): "How to Know Unknowable Things in Themselves"; in *Nous* 43, pp. 49-63

Hoppe, Hansgeorg (1983): *Synthesis bei Kant: Das Problem der Verbindung von Vorstellungen und ihrer Gegenstandsbeziehungen in der 'Kritik der reinen Vernunft'*; Walter De Gruyter

Hudson, Hud (1994): *Kant's Compatibilism*; Cornell University Press

Hume, David (1993): *An Enquiry Concerning Human Understanding*. Second Edition, edited by Eric Steinberg; Hackett Publishing Company

Hume, David (2009): *A Treatise of Human Nature*. Second Edition, with text revised and notes by P.H. Nidditch; Oxford University Press

Ilting, Karl-Heinz (1972): "Der naturalistische Fehlschluss bei Kant"; in Manfred Riedel, *Rehabilitierung der praktischen Philosophie, Band I: Geschichte, Probleme, Aufgaben*, pp. 79-97. Rombach

Irwin, Terence (1984): "Morality and Personality: Kant and Green"; in Allen Wood (ed.), *Self and Nature in Kant's Philosophy*, pp. 31-56. Cornell University Press

Irwin, Terence (2007): *The Development of Ethics*, Volume II. Oxford University Press

Johnson, Andrew B. (2005): "Kant's Empirical Hedonism"; in *Pacific Philosophical Quarterly* 86 (1), pp. 50-63

Joyce, Richard (2007): *The Myth of Morality*; Cambridge University Press

Kant, Immanuel (Anth): *Anthropology from a Pragmatic Point of View*; translated by Robert B. Loudon, Cambridge University Press, 2006

Kant, Immanuel (KU): *Critique of Judgment*; translated by J.C. Meredith, Oxford University Press, 1952

Kant, Immanuel (KpV): *Critique of Practical Reason*; translated by T.K. Abbott, Prometheus Books, 1996

Kant, Immanuel (CPR): *Critique of Pure Reason*; translated by Norman Kemp Smith, Palgrave Macmillan, 2003

Kant, Immanuel (EEUK): *First Introduction to the Critique of Judgment*; translated by Werner Pluhar, Hackett, 1989

Kant, Immanuel (GMS): *Groundwork for the Metaphysics of Morals*; translated by T. K. Abbott, Broadview Press, 2003

Kant, Immanuel (Log): *Lectures on Logic*; edited by J. Michael Young, Cambridge University Press, 2004

Kant, Immanuel (Man): *Metaphysical Foundations of Natural Science*; translated by Michael Friedman, Cambridge University Press, 2004

Kant, Immanuel (MM): *Metaphysics of Morals*; translated by Mary Gregor, Cambridge University Press, 2004

Kant, Immanuel (Prol): *Prolegomena to Any Future Metaphysics*; translated by Gary Hatfield, Cambridge University Press, 2004

Kant, Immanuel (RGV): *Religion Within The Limits Of Pure Reason*; translated by T. M. Greene and H.H. Hudson, Harper and Brothers, 1960

Kant, Immanuel (RezSchulz): *Review of Schulz's Attempt at an introduction to a doctrine of morals for all human beings regardless of different religions*; translated by Mary Gregor, *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*, Cambridge University Press, 1996

Kim, Jaegwon (2003): "The American Origins of Philosophical Naturalism"; in *Journal of Philosophical Research*, APA Centennial Volume, pp. 83-98

Kitcher, Patricia (1984): "Kant's Real Self"; in Allen Wood (ed.), *Self and Nature in Kant's Philosophy*, pp. 113-47. Cornell University Press

Kitcher, Patricia (1991): "Kant's Dedicated Cognitivism"; in J.C. Smith (ed.), *Historical Foundations of Cognitive Science*, pp. 189-210. Kluwer Academic Publishing

Kitcher, Patricia (1993): *Kant's Transcendental Psychology*; Oxford University Press

Kitcher, Patricia (1999): "Kant's Epistemological Problem and Its Coherent Solution"; in *Nous* 33, pp. 415-41

Kitcher, Philip (1986): "Projecting the Order of Nature"; in Robert E. Butts (ed.), *Kant's Philosophy of Physical Science*, pp. 201-35. Reidel

Kolodny, Niko (2005): "Why Be Rational?"; in *Mind* 114, pp. 509-63

Korsgaard, Christine (1996a): *Creating the Kingdom of Ends*; Cambridge University Press

Korsgaard, Christine (1996b): *The Sources of Normativity*; Cambridge University Press

Korsgaard, Christine (1997): "The Normativity of Instrumental Reason"; in Garrett Cullity, Beryt Gaut (eds.), *Ethics and Practical Reason*, pp. 215-54. Oxford University Press

Korsgaard, Christine (2009): *Self-Constitution: Agency, Identity, and Integrity*; Oxford University Press.

Kvist, Hans-Olof (1978): *Zum Verhältnis von Wissen und Glauben in der kritischen Philosophie Immanuel Kants*. Åbo akademi

Langton, Rae (1998): *Kantian Humility: Our Ignorance of Things in Themselves*; Oxford University Press

Lavin, Douglas (2004): "Practical Reason and the Possibility of Error"; in *Ethics* 114 (3), pp. 424-57

Lehrer, Keith (1990): *Theory of Knowledge*; Westview

Lewis, David (1973a): "Causation"; in *The Journal of Philosophy* 70, pp. 556-67

Lewis, David (1973b): *Counterfactuals*; Basil Blackwell

Lewis, David (1983): "New Work for a Theory of Universals"; in *Australasian Journal of Philosophy* 61, pp. 343-77

Lewis, David (1994): "Humean Supervenience DeBugged"; in *Mind* 103, pp. 473-90

Lewis, David (2000): "Dispositional theories of value"; in Lewis, *Papers in Ethics and Social Philosophy*, pp. 68-94. Cambridge University Press

Lewis, David (2003): "Are We Free to Break the Laws?"; in Gary Watson (ed.), *Free Will*, pp. 122-29. Oxford University Press

Locke, Don (1973-4): "Natural Powers and Human Abilities"; in *Proceedings of the Aristotelian Society* 74, pp. 171-87

Loewer, Barry (1996): "Freedom from Physics: Quantum Mechanics and Free Will"; in *Philosophical Topics* 24/2, pp. 91-112

Loewer, Barry (2004): "Humean Supervenience"; in John Carroll (ed.), *Readings on Laws of Nature*, pp. 176-206. University of Pittsburgh Press

Longuenesse, Beatrice (2001): *Kant and the Capacity to Judge*; Princeton University Press

Longuenesse, Beatrice (2009): *Kant on the Human Standpoint*; Cambridge University Press

Ludwig, Bernd (1999): "Warum es keine 'hypothetischen Imperative' gibt, und warum Kants hypothetisch-gebietende Imperative keine analytischen Sätze sind"; in H. F. Klemme, B. Ludwig, M. Pauen and W. Stark, *Aufklärung und Interpretation: Studien zu Kants Philosophie und ihrem Umkreis. Tagung aus Anlass des 60. Geburtstags von Reinhard Brandt*, pp. 105-24. Königshausen und Neumann

MacKay, D.M. (1960): "On the Logical Indeterminacy of a Free Choice"; in *Mind* 69, pp. 30-40

Mackie, J. L. (1991): *Ethics. Inventing Right and Wrong*; Penguin Books

Matthews, H.E. (1969): "Strawson on Transcendental Idealism"; in *Philosophical Quarterly* 19, pp. 204-20

McDowell, John (1978): "Are Moral Requirements Hypothetical Imperatives?"; in *Proceedings of the Aristotelian Society, Supplementary Volumes*, Vol. 52, pp. 13-29

McDowell, John (1994): *Mind and World*; Harvard University Press

McDowell, John (1998a): "Having the World in View: Sellars, Kant, and Intentionality"; in: *The Journal of Philosophy* 1998, XCV (9), pp. 431-91

McDowell, John (1998b): "Might There Be External Reasons?"; in McDowell, *Mind, Value and Reality*, pp. 95-111. Harvard University Press

Meerbote, Ralf (1982): "Wille and Willkür in Kant's Theory of Action"; in Moltke S. Gram (ed.), *Interpreting Kant*, pp. 69-89. University of Iowa Press.

Meerbote, Ralf (1984a): "Kant on the Nondeterminate Character of Human Actions"; in W. L. Harper and Meerbote (eds.), *Kant on Causality, Freedom, and Objectivity*, pp. 138-163. University of Minnesota Press

Meerbote, Ralf (1984b): "Kant on Freedom and the Rational and Morally Good Will"; in Allen Wood (ed.), *Self and Nature in Kant's Philosophy*, pp. 57-72. Cornell University Press

Meerbote, Ralf (1991): "Kant's Functionalism"; in J.C. Smith (ed.), *Historical Foundations of Cognitive Science*, pp. 161-88. Kluwer Academic Publishing

Melnick, Arthur (1973): *Kant's Analogies of Experience*; University of Chicago Press

Milgram, Elijah (1997): *Practical Induction*; Harvard University Press

- Milican, Peter, J.C. (2002): "Hume's Skeptical Doubts Concerning Induction"; in Milican (ed.), *Hume on Human Understanding*, pp. 107-73. Clarendon Press
- Moore, G.E. (1912): *Ethics*; Oxford, U.K.: Oxford University Press
- Moran, Richard (2001): *Authority and Estrangement. An Essay on Self-Knowledge*. Princeton University Press
- Nagel, Thomas (1989): *The View From Nowhere*; Oxford University Press
- Nelkin, Dana (2000): "Two Standpoints and the Belief in Freedom"; in *The Journal of Philosophy* 97 (10), pp. 564-76
- Nietzsche, Friedrich (1966): *Beyond Good and Evil*, translated by Walter Kaufmann; Random House
- Nozick, Robert (1977): *Anarchy, State and Utopia*; New York: Basic Books.
- O'Neill, Onora (1989): "Reason and autonomy in Grundlegung III"; in O'Neill, *Constructions of Reason*, pp. 51-65. Cambridge University Press
- Owens, David (2000): *Reason without Freedom. The problem of epistemic normativity*; Routledge
- Papineau, David (1992): "Reliabilism, Induction, and Skepticism"; in *The Philosophical Quarterly* Vol. 42, No. 166, pp. 1-20
- Parfit, Derek (1986): *Reasons and Persons*; Oxford University Press
- Parfit, Derek (2011): *On What Matters* (Volume 1, 2); Oxford University Press
- Paton, H.J. (1958): *The Categorical Imperative. A Study in Kant's Moral Philosophy*; Hutchinson & Co.
- Paton, H.J. (1970): *Kant's Metaphysics of Experience*; Allen and Unwin
- Pereboom, Derek (2005): "Kant on Transcendental Freedom"; in *Philosophy and Phenomenological Research* 73 (3), pp. 537-67
- Perry, David (1965): "Prediction, Explanation, and Freedom"; in *Monist* 49, no. 2
- Pettit, Philip and Smith, Michael (2003): "Freedom in Belief and Desire"; in: Gary Watson (ed.), *Free Will*, pp. 388-407. Oxford University Press

- Pippin, Robert (1987): "Kant on the Spontaneity of Mind"; in *Canadian Journal of Philosophy* 17, pp. 449-75
- Pogge, Thomas (1991): "Erscheinungen und Dinge an sich"; in: *Zeitschrift fuer Philosophische Forschung* Band 45, Heft 4, pp. 489-510
- Pollok, Konstantin (2007): "'Wenn Vernunft Volle Gewalt Über das Begehungsvermögen Hätte' - Über Die Gemeinsame Wurzel der Kantischen Imperative"; in *Kant-Studien* 98 (1), pp. 57-80
- Popper, Karl (1950): "Indeterminism in Quantum Physics and in Classical Physics. Part I"; in *British Journal for the Philosophy of Science* 1, pp. 117-33
- Prauss, Gerold (1971): *Erscheinung bei Kant*; De Gruyter
- Prauss, Gerold (1974): *Kant und das Problem der Dinge an sich*; Bouvier.
- Prauss, Gerold (1983): *Kant über Freiheit als Autonomie*; Vittorio Klostermann
- Price, Anthony (2009): "Slaves of the Passions By Mark Schroeder" (Review); in *Philosophy* 84 (2), pp. 291-5
- Quine, W.V. (1951): "Two Dogmas of Empiricism"; in *The Philosophical Review* 60, pp. 20-43
- Quine, W.V. (1986): "Reply to Jules Vuillemin"; in Lewin Hahn and Paul Schilpp (eds.), *The Philosophy of W.V. Quine*. Open Court
- Quinn, Warren (1994): *Morality and Action*; Cambridge University Press
- Railton, Peter (1984): "Alienation, Consequentialism, and the Demands of Morality"; in *Philosophy and Public Affairs* 13, pp. 134-171
- Railton, Peter (1986): "Moral Realism"; in *The Philosophical Review* 95, pp. 163-207
- Railton, Peter (1997): "On the Hypothetical and Non-Hypothetical in Reasoning about Thought and Action"; in Garrett Cullity, Beryt Gaut (eds.), *Ethics and Practical Reason*, pp. 53-80. Oxford University Press
- Rawls, John (1989): "Themes in Kant's Moral Philosophy"; in E. Förster (ed.), *Kant's Transcendental Deductions*, pp. 81-113. Stanford University Press
- Raz, Joseph (2010): "Reason, Reasons and Normativity"; in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Volume 5, pp. 5-24. Oxford University Press

Reath, Andrews (1989): "Hedonism, Heteronomy, and Kant's Principle of Happiness"; in *Pacific Philosophical Quarterly* 70, pp. 42-72

Reath, Andrews (2006): *Agency and Autonomy*; Oxford University Press

Reibenschuh, Gernot (1990): "Über das 'Faktum der reinen Vernunft'. Eine Skizze."; in G. Funke (ed.), *Akten des Siebten Internationalen Kant-Kongresses*, pp. 365-74

Rush, Fred L. Jr. (2000): "Reason and Regulation in Kant"; in *Review of Metaphysics* vol. 53, pp. 837-62

Ryle, Gilbert (1949): *The Concept of Mind*; Hutchinson's University Library

Robinson, Hauke (1994): "Two Perspectives on Kant's Appearances and Things in Themselves"; in *Journal of the History of Philosophy* 32, pp. 411-41

Russell, Bertrand (1992): "On the Notion of Cause"; in J. Slater (ed.), *The Collected Papers of Bertrand Russell v6: Logical and Philosophical Papers 1909-1913*, pp. 193-210. Routledge

Salmon, Wesley (1984): *Scientific Explanation and the Causal Structure of the World*; Princeton University Press

Scanlon, Thomas (2000): *What We Owe To Each Other*; Harvard University Press

Scanlon, Thomas (2003): "The Significance of Choice"; in: Gary Watson (ed.), *Free Will*, pp. 352-71. Oxford University Press

Schaffer, Jonathan (2001): "Causes as Probability-Raisers of Processes"; in *The Journal of Philosophy* 98 (2), pp. 75-92

Schapiro, Tamar (2009): "The Nature of Inclination"; in *Ethics* 119, pp. 229-56

Scheffler, Samuel (2010): "Valuing"; in Scheffler, *Equality and Tradition: Questions of Value in Moral and Political Philosophy*, pp. 15-40. Oxford University Press

Schönecker, Dieter (1999): *Kant: Grundlegung III. Die Deduktion des kategorischen Imperativs*. Karl Alber

Schönecker, Dieter/Wood, Allen (2002): *Kants 'Grundlegung zur Metaphysik der Sitten'. Ein einführender Kommentar*. Schöningh

Schönecker, Dieter (2005): *Kants Begriff transzendentaler und praktischer Freiheit. Eine entwicklungsgeschichtliche Studie*. Walter de Gruyter

- Schroeder, Mark (2007): *Slaves of the Passions*; Oxford University Press
- Scriven, Michael (1968): "An Essential Unpredictability in Human Behavior"; in B. Wolman and E. Nagel (eds.), *Scientific Psychology: Principles and Approaches*. Basic Books
- Senderowicz, Yaron (2005): *The Coherence of Kant's Transcendental Idealism*; Springer, 2005
- Sellars, Wilfrid (1970): "...this I or he or it (the thing) which thinks..."; in *Proceedings of the American Philosophical Association* 44, pp. 5-31
- Shafer-Landau, Russ (2005): *Moral Realism: A Defence*; Oxford University Press
- Shah, Nishi (2003): "How Truth Governs Belief"; in *The Philosophical Review*, Vol. 112, pp. 447-82
- Sider, Theodore (1992): "Tooley's Solution to the Inference Problem"; in *Philosophical Studies* 67, pp. 261-75
- Silber, John (1963): "The Highest Good in Kant's Ethics"; in: *Ethics* 73, pp. 179-97
- Smart, J.J.C. (2003): "Free Will, Praise, and Blame" in Gary Watson (ed.), *Free Will*, pp. 58-71. Oxford University Press
- Smith, Michael (1994): *The Moral Problem*; Blackwell
- De Sousa, Ronald (1984): "The Natural Shiftiness of Natural Kinds"; in *Canadian Journal of Philosophy* Vol. 14, No. 4, pp. 561-80
- Steigleder, Klaus (2002): *Kants Moralphilosophie. Die Selbstbezüglichkeit reiner praktischer Vernunft*; Metzler
- Strawson, Galen (2004): "Mental ballistics or the involuntariness of spontaneity"; in *Proceedings of the Aristotelian Society Volume* 103 (1), pp. 227-56
- Strawson, P. E. (1966): *The Bounds of Sense. An Essay on Kant's Critique of Pure Reason*; Methuen
- Strawson, P. E. (1987): *Skepticism and Naturalism: Some Varieties*; Columbia University Press
- Strawson, P. E. (2003): "Freedom and Resentment"; in Gary Watson (ed.), *Free Will*, pp. 72-93. Oxford University Press
- Stroud, Barry (1999): *The Quest for Reality. Subjectivism and the Metaphysics of Color*; Oxford University Press

- Stroud, Barry (2006): "The Constraints of Hume's Naturalism"; in *Synthese* 2006, pp. 339-51
- Stroud, Barry (2011): *Engagement and Metaphysical Dissatisfaction: Modality and Value*; Oxford University Press
- Taylor, Charles (1985): "What is Human Agency?"; in Taylor, *Human Agency and Language: Philosophical Papers I*, pp. 15-44. Cambridge University Press.
- Tooley, Michael (1987): *Causation: A Realist Approach*; Oxford University Press
- Tooley, Michael/Sosa, Ernest: "Introduction"; in in Tooley/Sosa (eds.), *Causation*, pp. 1-32. Oxford University Press
- Vaihinger, Hans (1949): *The Philosophy of 'As if...'*; Routledge and Keagan
- Van Cleve, James (1999): *Problems from Kant*; Oxford University Press
- Van Inwagen, Peter (1983): *An Essay on Free Will*; Clarendon Press
- Velleman, David (2000): *The Possibility of Practical Reason*; Oxford University Press
- Walker, Ralph (1978): *Kant. The Arguments of the Philosophers*; Kegan and Paul
- Walker, Ralph (1983): "Empirical Realism and Transcendental Anti-Realism: Part II"; in *Proceedings of the Aristotelian Society*, Supplementary 57, pp. 155-77
- Walker, Ralph (1984): "Synthesis and Transcendental Idealism"; in: *Kant-Studien* 76, pp. 14-27
- Wallace, R. Jay (1994): *Responsibility and the Moral Sentiments*; Harvard University Press
- Wallace, R. Jay (2006): "Three Conceptions of Rational Agency"; in Wallace, *Normativity and the Will*, pp. 43-62. Clarendon Press
- Warren, Daniel (1998): "Kant and the Apriority of Space"; in *The Philosophical Review* 107, pp. 179-224
- Warren, Daniel (2001): *Reality and Impenetrability in Kant's Philosophy of Nature*; Routledge
- Warren, Daniel (unpublished): "Kant on Inner Determinations and Things in Themselves"
- Wartenberg, Thomas E. (1992): "Reason and the practice of science"; in Paul Guyer (ed.), *The Cambridge Companion to Kant*, pp. 228-48. Cambridge University Press
- Watkins, Eric (2004): *Kant and the Metaphysics of Causality*. Cambridge University Press

- Watson, Gary (2003): "Free Agency"; in Watson (ed.), *Free Will*, pp. 337-51. Oxford University Press
- Wiggins, David (1969): "Freedom, Knowledge, Belief and Causality"; in *The Royal Institute of Philosophy Lectures*, 3, pp. 132-54
- Wiggins, David (2003): "Towards a Reasonable Libertarianism"; in Gary Watson (ed.), *Free Will*, pp. 94-121. Oxford University Press
- Widerker, David (2003): "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities"; in Gary Watson (ed.), *Free Will*, pp. 177-89. Oxford University Press
- Willaschek, Marcus (1991): "Die Tat der Vernunft. Zur Bedeutung der Kantischen These vom 'Faktum der Vernunft'"; in: G. Funke (ed.), *Akten des Siebten Internationalen Kant-Kongresses*, pp. 455-66
- Willaschek, Marcus (1992): *Praktische Vernunft. Handlungstheorie und Moralbegründung*; J.B. Metzler
- Willaschek, Marcus (1998): "Phaenomena/Noumena und die Amphibolie der Reflexionsbegriffe"; in: G. Mohr, Willaschek (eds.), *Klassiker Auslegen: Immanuel Kant, Kritik der reinen Vernunft*, pp. 325-51. Akademie Verlag
- Williams, Bernard (1973) "Deciding to believe"; in Williams, *Problems of the Self*, pp. 136-51, Cambridge University Press
- Williams, Bernard (1986a): *Ethics And The Limits of Philosophy*; Harvard University Press
- Williams, Bernard (1995): "How free does the will need to be?" In: Williams, *Making Sense of Humanity*; Oxford University Press
- Williams, Bernard (1997): "Internal and External Reasons"; in S. Darwall, A. Gibbard, P. Railton (eds.), *Moral Discourse and Practice*, pp. 363-72. Oxford University Press
- Williams, T. C. (1968): *The Concept of the Categorical Imperative*; Clarendon Press
- Wolf, Susan (1980): "Asymmetrical Freedom"; in *The Journal of Philosophy* 77, pp. 151-66
- Wolf, Susan (1990): *Freedom within Reason*; Oxford University Press
- Wolf, Susan (2003): "Sanity and the Metaphysics of Responsibility"; in Gary Watson (ed.), *Free Will*, pp. 372-87. Oxford University Press

Wolff, Robert Paul (1963): *Kant's Theory of Mental Activity: a commentary on the transcendental analytic of the Critique of Pure Reason*; Harvard University Press

Wood, Allen (1984): "Kant's Compatibilism"; in Wood (ed.), *Self and Nature in Kant's Philosophy*, pp. 73-101. Cornell University Press

Woodward, James (1990): "Laws and Causes"; in *The British Journal for the Philosophy of Science* 41, pp. 553-73

Woodward, James (2003): *Making Things Happen: A Theory of Causal Explanation*; Oxford University Press