Title
Technical Warrant and its Relation to Intentional Action

Author
Iterum, Skef

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Technical Warrant and its Relation to Intentional Action

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Philosophy

by

Skef Iterum

2017

ABSTRACT OF THE DISSERTATION

Technical Warrant and its
Relation to Intentional Action

by

Skef Iterum
Doctor of Philosophy in Philosophy
University of California, Los Angeles, 2017
Professor Alexander Jacob Julius, Chair

We are often highly confident that we can bring about what we intend, but current theories of intention only partly account for this capacity. The theory of epistemic warrant can serve as a model for a corresponding theory of *technical warrant* for appropriately confident intention: What permits confidence in an *aim* to perform a "basic action" under a certain condition is a *technical entitlement* corresponding to the epistemic entitlement to a perceptual belief. Basic aims, together with beliefs, can then *technically justify* non-basic aims, not as valuable or desirable things to do, but as ends pursued in virtue of the justifying means.

In light of the theory of technical warrant, knowledge itself provides a model for a theory of intentional actions as technically warranted true aims in the absence of "waywardness": What has been discussed as "antecedential" waywardness in the context of Causalist theories of action is analogous to veridical hallucination. Similarly, problems of "consequential" and "tertiary" waywardness correspond directly to Gettier problems, in that each is the result of a "gap" between a specific pattern of justificatory reasoning and circumstances in the world. Intentions of lower confidence are analogous to lower-credence epistemic attitudes.

The dissertation of Skef Iterum is approved.

Andrew Hsu

Adnan Youssef Darwiche

Calvin G. Normore

C. T. Burge

Alexander Jacob Julius, Committee Chair

University of California, Los Angeles

2017

For Scott and Gabriel,

to whom I am unusually close

TABLE OF CONTENTS

VITA

1994	Bachelor's Degree with Honor in Engineering and Applied Science, California Institute of Technology.
1996	Master's Degree in Computer Sciences, University of Wisconsin, Madison.
2000–2010	Principle Investigator on five patents (seven total)
2007–2009	Special Student, University of Wisconsin, Madison.
2011	Masters Degree, University of California, Los Angeles.

PUBLICATIONS

"Inheritance-Inspired Interface Versioning for CORBA" (with R. Campbell). *ECOOP 2004—Object-Oriented Programming, 18th European Conference, Oslo, Norway.*

CHAPTER 1

Introduction

1.1 Topic and Scope

The subject of this thesis is the nature of intentional action. Our intentions are, speaking roughly and broadly, our plans—what we hope to bring about in the world. And our intentional actions are more or less the successful realizations of those plans. I decide to make a peanut butter sandwich, and thereby form an intention to make one. My spreading peanut butter on a slice of bread is part of my making the sandwich, and also something I do intentionally in its own right. If I perform additional sufficient steps, such as placing another slice on that one, and certain other tricky things go right, I make the sandwich intentionally.

The questions motivating this work concern only part of the more general subject of *action*. If my actions are those things that can correctly be characterized as what I *do*, then my tripping on a carpet or blinking my eye are usually not intentional. *Unintentional* actions are not a focus of this thesis. There are also aspects of intentional action I do not explore in any depth, such as the cases explored in Knobe (2003), in which others ascribe intentional actions to a person who does not have (or at least does not take herself as having) a corresponding intention. I do not address the question of free will, and have relatively little to say about will or volition in general[1]. However, as others have noted[2], the nature of intention and its relation to intentional action are central to action theory. Knobe's "side effects" are closely related to intentions of the agent in question, and the scope of what one does is somewhat tied to the scope of what

1. What I do say on this subject is in Section 6.6.1.

2. See, for example, Audi (1973, p. 401–2) and Mele (2003, §4).

one can do intentionally; I blink my eyes unintentionally but *could* intend to blink them, while I do not beat my heart and—not being a Buddhist adept—I cannot intend that it beat (at least by any normal means).

Because my interest is in the degree to which the theory of Chapters 6, 7 and 10 can account for the phenomenon of intentional action, I focus more on how it addresses proposed norms of intention and key cases than on how it compares to other theories. The exception to this rule relates to a problem that has not been of primary concern in recent action theory. This is our capacity to act with a degree of confidence appropriate to our skills and warranted beliefs about our situation, and more particularly to have an appropriately *high* degree of confidence. This capacity is the subject of Chapter 2, and I discuss the extent to which other theories account for appropriate confidence in Chapter 3.

I focus on this capacity because accounting for it is a primary motivation for the thesis as a whole. Whether or not our possessing this capacity is the most important or central aspect of our agency, I have come to think that it is the most pervasive and "structuring" aspect: more than anything else, it determines the form of our intentions and how we either act on or revise them. Accordingly, it poses the most significant obstacle in the path of theories of intention that do not, or at least do not yet, account for it. This attitude does not appear to be widely shared; while appropriate confidence in intention is often discussed, it is usually treated as a secondary concern at best. The reason for this relative neglect appears to trace to a lack of primary action *concept* to tie it to. I discuss this issue in Sections 2.3 and 2.4.

1.2 Knowledge and Intentional Action as Inverses

So what is the theory in a nutshell? Current standards of scholarship call for an overview of what is to come, if not now then earlier. Unfortunately, not every theory can be stated in a way that is concise, lucid, and satisfying. If there is such a summary of this work, I have not yet arrived at it. So I will begin with a concise statement and, in the remainder of this introduction, will do my best to offer a lucid explanation. That will prepare the way for the

remaining chapters and, I hope, some eventual satisfaction.

In it's most concise form, my thesis is that:

> Intentional action is best understood as the inverse of knowledge.

Right off, this needs some "unpacking", starting with the weasel-words "is best understood as". Will I argue that intentional action is the inverse of knowledge, or not? They are certainly not complete mathematical or computational inverses, given that it is not possible for one to *do* anything that one can *know*, such as things in the past. Even so, mathematics, at least, also provides for *partial* inverse functions, and therefore we could treat each as the inverse of the other under certain restrictions of domain. However, because I will draw on both conceptions of *inverse*, and there are anyway some other important distinctions between the two phenomena, the relationship can be thought of as metaphorical.

As a metaphor, though, the thesis serves as a fruitful answer to much-discussed question "What is the relation of knowledge to intentional action?". The motivation for that question is the shared sense that there is some relationship between the two, and that understanding it would shed light on the nature of intention and action. The thesis suggests an investigative strategy: To best understand intentional action, we should look to our understanding of knowledge—that is, to epistemology—and consider how one might "turn it around", so that what were its "outputs" become inputs, and vice-versa. Implicit in this strategy is that we have a better grasp on the nature of knowledge than we do of intentional action, which is my view and a premise of this work. The result of this investigation could be accurately called an "inverse theory" of action, given the history of its construction.

So what, then, are the "inputs" and "outputs" of knowledge, and in what sense might one try to "turn it around"?

1.2.1 The Mathematical Metaphor of an Inverse Function

Consider the metaphor at its simplest level, in virtue the concept of a mathematical function and its inverse. A mathematical function is just a mapping from a "domain" to a "range". In

the case of knowledge, we might think of the domain as states of affairs (however conceived) and the range as beliefs, each of those being a proposition (however conceived) towards which we take the *attitude* that it is true. Even saying so little, this arrangement is already not quite right, as more than one belief can represent the same state of affairs. That problem can stand as the first of many approximations. In the corresponding picture of intentional action, the domain would be what we might temporarily call "goals", considered as propositions that we take a different attitude towards—roughly of *wanting it to come about*—with the range being states of affairs.

In this most basic application of the concept of inverse, the metaphor captures no more than the familiar idea of "direction of fit": Is the proposition somehow held to the standard of the state of affairs, or is the state of affairs somehow held to the standard of the proposition? In fact, "captures" may understate the relation between the two ideas. What is this so natural-sounding "direction", after all? It is clearly not one in space, and not plausibly one in time—we have at least some knowledge of the future. Instead, the idea seems to be that the fit goes from the proposition to the world in one case, and from the world to the proposition—the *opposite direction*—in the other. This is already something like *fit* being "turned around".

The idea of direction of fit can only reveal so much about the relationship between knowledge and intentional action, however, given its generality. After all, the canonical attitude with world-to-mind fit is not intention but *desire*, usually conceived as a representation of something that one wants but may make no attempt to get. For the metaphor of the inverse relationship to tell us more, it needs to be more specific. And the best route I have found to a more detailed understanding is through Anscombe's example of the shopper and the detective. That example, which is likely the most famous part of *Intention*, goes as follows (Anscombe 1957, §32):

> Let us consider a man going round a town with a shopping list in his hand. Now it is clear that the relation of this list to the things he actually buys is one and the same whether his wife gave him the list or it is his own list; and that there is a different relation when a list is made by a detective following him about. If he made the list himself, it was an expression of intention; if his wife gave it him, it has the role of an order. What then is the identical relation to what happens, in

> the order and the intention, which is not shared by the record? It is precisely this: if the list and the things that the man actually buys do not agree, and if this and this alone constitutes a *mistake*, then the mistake is not in the list but in the man's performance (if his wife were to say 'Look, it says butter and you have bought margarine', he would hardly reply: 'What a mistake! I must put that right' and alter the word on the list to 'margarine'); whereas if the detective's record and what the man actually buys do not agree, then the mistake is to the record.

This paragraph, or a summary of it, has probably been most widely *encountered* as an excellent example of a philosophical example, which is how it appears in her Guardian obituary (O'Grady 2001). But it is most widely *cited* in philosophical contexts as an illustration of the concept of direction of fit. The example was first tied to that concept by John Searle, who called it "the best illustration of [the difference in the direction of fit between words and the world] I know of" (Searle 1975, p. 3). The association is now so strong that it is sometimes attributed to Anscombe herself, although most citations at least acknowledge that she doesn't use that *term*.

These two facts sit awkwardly together because the concept of direction of fit is pretty boring. Consider the case of "The Day-dreamer and the Detective":

> The day-dreamer wants a doughnut, a desire that prompts him to doodle one in the margins of his homework. The detective is hired to find out what, if anything, the day-dreamer receives that day. In a coincidence, the day-dreamer's roommate happens to come home with a doughnut and gives it to him, fulfilling his desire. The attentive detective, looking through the window, writes down that the day-dreamer received a doughnut.

This description provides a perfectly serviceable contrast between directions of fit, but I doubt that even Anscombe's prose style could transform it into one of philosophy's famous cases.

Indeed, little that Anscombe says would apply to the ordinary conception of desire as distinct from intention. To have a desire be realized is nice—most of the time?—but not having it realized does not mean that it has failed as a desire, or really that anything has "gone wrong" in a normative sense. It can be more of a mistake, of an admittedly somewhat different kind, to act on a desire than to avoid acting on it. And although there are often referential relations between the content of a realized desire and objects or facts in the world that constitute its

realization, it is not clear that there is any principled way of associating an unrealized desire with a particular object or fact that *failed* to realize it, unless the desire was also an intention. So while the passage does evoke the concept that Searle attributes to it[3], it also evokes much more.

What makes Anscombe's example *interesting* are the similarities between the shopper's list and his shopping on the one hand, and the detective's list and her detecting on the other. The shopper's list, of which Anscombe says "if he made the list himself, it is an expression of intention", bears, or fails to bear, a certain relationship to what he does about town. Anscombe compares this relationship to that between the detective's list (which might be called an "expression of belief") and what *she* does about town. If the shopping goes well, the shopper has bought some items intentionally; if the detecting goes well, the detective knows what items were bought. Each of these is a kind of success of "fit" at least partly attributable to the person involved, in a way that the simple realization of a desire need not be.

At the same time, consistent with but going beyond the issue of direction of fit, there are other aspects of the example that are the reverse of one other. With the detecting, there is first some observation and activity, and the list is a product of that. If what has been bought is not what is recorded, something has gone wrong in the creation of the list. In the case of the action, the list sets a kind of standard for the shopping, so that buying fewer or different things would constitute a "mistake of performance". As Anscombe clarifies in the next paragraph, however, this standard is subject to "qualification":

> In the case of a discrepancy between the shopping list and what the man buys, I have to introduce the qualification: If this and this alone constitutes a mistake. For the discrepancy might arise because some of the things were not to be had and if one might have known they were not to be had, we might speak of a mistake (an error of judgment) in constructing the list. If I got out in Oxford with a shopping list including 'tackle for catching sharks', no one will think of it as a mistake in performance that I fail to come back with it. And then again there may be a discrepancy between the list and what the man bought because he changed his mind and decided to buy something else instead.

3. However, anyone interested in Anscombe's take on the questions that concern Searle are likely to find her discussion of intentions and predictions in Anscombe (1957, §2) more directly pertinent.

Each of these two pre-conditions play an important role in the chapters to come. I will discuss the second, more familiar one, first. Both what we intend and what we believe are, to a certain extent, up to us. They are not *entirely* up to us, of course. One should not, and probably cannot, believe *anything* against the evidence, and it is at least doubtful that one can intend *anything*. Still, many of our beliefs are voluntary in the narrow sense that we can choose what questions to spend time and effort answering. Still, one way that intentions and beliefs differ is that once you arrive at a belief, it is either something you know or it isn't[4]. Your belief is the culmination of a process that you are at least partly responsible for. An intention, in contrast, is not the culmination of its associated process; it comes earlier. And, importantly, an intention that is revoked in response to a change of motivation or circumstance *does not succeed or fail.*

This feature of intention does not obviously generalize across other attitudes with the same direction of fit. If the day-dreamer hopes to receive a doughnut at three this afternoon, and then loses that desire in eating a late lunch, his subsequently receiving a doughnut at three will still technically realize that desire. It is "satisfied" in Searle's philosophical sense, whether or not the day-dreamer is in the ordinary one[5]. For present purposes, however, the important distinction is that at the point one is arrived at, an intention has a different status than a belief; a belief already has its status in the larger epistemic context, while an intention is provisional. A theory of intention must therefore account for this distinction, which poses a difficulty for *Cognitivist* theories of action, according to which an intention is a variety of belief.

Rather than picking out a key distinction between knowledge and intentional action, Anscombe's first pre-condition picks out a key similarity—one that the first half of this thesis is largely structured around, and the second half largely explores the implications of. For failed shopping to be a failure of *performance*, Anscombe says that the items on the list must not be what "one might have known [were] not to be had". The importance of this requirement is in how it carves up the "space" of intentional action. It is implausible to read Anscombe here

4. Some may object that beliefs about the future aren't knowledge until the facts are settled, but that worry is beside this point, because once the relevant facts are settled, we can look back and ask "Did she know then?"

5. Searle himself accounts for this difference within his "conditions of satisfaction" framework. I briefly argue against this approach in Section 7.1.

as ruling out *trying* to buy tackle for catching sharks in Oxford under these circumstances. Therefore, what puts an intended end in the grasp of one's good performance, in her view, is a certain kind of judgment in arriving at that end, which is based on what "one might have known" about the circumstances in which one will eventually act. While this is more of an observation than an explanation, it does imply that judgment about your circumstances is part of what allows an action to be up to you.

This is one of only a few parts of *Intention* that draws a explicit line to the high degree of confidence in "doing" as opposed to the lower confidence in what I will temporarily characterize as "trying". Other parts, such as her view that we have "knowledge without observation" of what we do, might read differently in that light. However, my object here is not to argue for a particular interpretation of *Intention*, but to suggest what we might take away from §32. If we do intend[6] with varying degrees of (generally) appropriate confidence, the cases of intentional action most comparable to cases of knowledge will be those in which we are *highly* confident of our ability to bring about our intention.

With that in mind, how might we restrict our two metaphorical functions to instances of high confidence? The knowledge function already comes close, in that belief is specifically an attitude of high credence. To be knowledge, however, the high credence of belief must be *appropriate*, in virtue of being epistemically *justified* or, more generally, epistemically *warranted*[7]. We might therefore add warrant as a parameter of the function. States of affairs will now map not just to beliefs, but to belief/warrant pairs; with the same belief potentially having various different justifications.

The intentional action function needs two changes. The first is that the attitude towards the goal must be, like belief, one of *high* confidence of the sort alluded to by Anscombe. This is, again, not confidence that the goal is *right*, in that it represents the best choice in terms of ethics or value. It is confidence only that it is in one's power to bring about the goal. The second change is to add something to the domain of the function analogous to epistemic warrant, to

6. Using that term loosely, for reasons I ultimately discuss in Section 7.6.

7. Epistemic warrant is the subject of Chapter 5.

ensure that the agent's confidence that she can bring about the goal is well-placed.

The first half of this thesis is largely devoted to developing a theory of this "parameter". The details of that theory mirror those of a theory of epistemic warrant, but "restructured" to support a goal whose realization is up to the agent, instead of a belief, which is not. I argue that the "glue" of justification is what relates means to ends. This distinctive type of warrant is arrived at in a "realm" of *technical* reasoning distinct from the *epistemic* realm, which constitutes a separable subset of *practical* reasoning[8].

Even with these additions, however, neither function is quite correct. With respect to knowledge, Anscombe, writing some years before Gettier's famous paper, does not consider whether good detecting and an accurate list would always constitute knowledge. Following the now-conventional view that they might not, we could add a further parameter to the domain of the knowledge function that captures whatever complicated states of affairs preclude a Gettier problem. These states would (in a loose sense) be determined by the corresponding warrant, and for a given justification there would be many such arrangements. With this last addition, the function matches the view that these four elements: belief, epistemic warrant, the state of affairs the belief represents, and a (complicated) state of affairs that avoids a Gettier problem, can differentiate what is knowledge from what is not.

In the example of the shopper and the detective, at least, Anscombe also does not consider a possible "accidental realization" of a goal[9]. If the specification of a goal is as simple as the list of items that the shopper carries, it could be realized by the detective dropping off those items before the former even leaves the house, rendering his "performance" redundant. That sort of possibility is what Searle hopes to preclude in adding a bit of self-referential content to every intention. And it is what Causalists hope to preclude by stipulating that a realization is only intentional if it comes about in not merely the causal wake of the *agent*, but that of her *inten-*

8. I explain this terminology in Section 2.2.

9. She does raise the subject at least once: "For if after his saying 'To replenish the water-supply' we can say 'He is replenishing the water-supply', then this would, in ordinary circumstances, of itself be enough to characterize *that* as an intentional action. (The qualification is necessary because an intended effect just occasionally comes about by accident)." (Anscombe 1957, §23).

tion, considered as a state physically encoded in her brain. Even granting these stipulations, however, there is a widely-discussed body of cases of "causal waywardness" or "deviance" in which a goal is realized and the conditions are met, but the result does not seem genuinely intentional[10].

So what parameter might we add to the intentional action function to preclude waywardness? In Chapter 10 I argue that what is needed are, in effect, states of affairs directly analogous to those the preclude a Gettier problem, and that these are needed for much the same reason. The theory that Gettier problems and waywardness are related somehow is not new. As I discuss in more detail in Section 10.2, many have noted that each sort of problem bears a revealing relationship to the concept *accident*. There have also been proposals according to which knowledge of one's circumstances, as opposed to mere warranted beliefs, might preclude the possibility of waywardness.

In this thesis I make a more specific claim, which is not just that warrant for intention *resembles* epistemic warrant, or that it plays an analogous role. Rather, it is that the "internal structure" of each is the *same*. Once this shared aspect of intention and epistemic warrant is properly understood, problems of waywardness and Gettier problems can then be understood as the same family of problem as seen in technical and epistemic realms of reasoning, respectively. This similarity in structure of warrant and "gap" ultimately stems from the similarity in the structure of the problems each needs to overcome. However, to discuss such problems I must move beyond the mathematical notion of a function, which is just a mapping from domain to range, and therefore provides no context for "problems" and their "solutions".

1.2.2 Computation (Actual and Metaphorical)

In the process of extending the inverse relationship in the last section, I put epistemic warrant in the range and technical warrant in the domain of their respective functions. That arrange-

10. Some theories add a clause like Searle's that amounts to "if everything goes right otherwise". However likely it is that we do qualify our goals to ourselves this way, such theories still face the problem of specifying what needs to go right in a particular case.

ment may be appropriate in the metaphor, but I can offer no good independent argument for it. A justification does not seem to be strictly a starting or end point. The knowledge function, for example, may be sufficient for specifying instances of knowledge on the part of an individual, but this reveals nothing about what an epistemic warrant might be needed for—about the role that it plays.

However, we do understand what that role is, for the most part. A person has a limited amount of information about the world, and epistemic reasoning is a method of extending that information, or perhaps "rearranging" it, into a form that may be more useful. A given extension may not be perfect—there is a chance that a belief arrived at through reasoning will be false. Still, the technique works well in the general case, and the extended or rearranged information is sometimes very useful. An epistemic justification is a kind of record of the reasoning that supports one's belief[11]. That record allows one to communicate one's reasoning to others, so that not all communication of belief needs to rest entirely on trust, and to revise beliefs in light of new evidence without having to start entirely from scratch. One can imagine a creature whose beliefs (or "beliefs") are arrived at through good reasoning but never justified, because it does not generally remember the steps it took along the way. Over time, however, we would expect more of its beliefs to be false than ours.

Given these roles for justification, we might think metaphorically of an "epistemic reasoning machine" that starts with some justified beliefs and arrives at new ones that are likely to be true, while keeping a record of the steps it took along the way. When it arrives at a contradiction, the machine can use its record to trace back through its past reasoning and make adjustments. To the extent that beliefs can be associated with information, what this machine does is "process information", which makes it tempting to refer to it as a computer. I believe this metaphor can be fruitful, but can also lead to misunderstanding, and raise concerns about how much the arguments of this thesis depend on the so-called "computational theory of mind". My concise answer to those concerns is "I think not too much", but this claim also needs some unpacking.

11. Or perhaps, in some cases, of reasoning that could support it just as well.

In the chapters that follow I often make use of what might be called "computational theory of *brain*" when considering the nature of *unconscious* mental processes. An example: We have some idea of what parts of the brain are responsible for visual perception, because damage to those parts results in problems of vision as opposed to other problems. Research on particular damages of a finer grain, and in virtue of other technologies like EEG and MRI, provides some insight into the functions of smaller blobs of brain[12], and what information the neural signals going into and out of a blob may correspond to (which makes either the computational or teleological sense of "function" applicable). The terminology of computation is convenient for considering how encoded information is transformed inside a blob, or modeling how many different blobs might function together.

Conscious thought processes raise two additional kinds of concern in relation to computation. One is that a computer might be simply unable to do what a conscious mind does, even if the problem is limited to one in which sensory data are the "inputs", and bodily movements (including those that produce speech) the "outputs", along with a lot of additional state to serve as a "memory". In that case, some of our cognitive capacities could not even be accurately simulated on a computer. The other concern is that even if a computer could behave as a person does, it would lack something important that we have. Perhaps it would have no experiences, or the information it processed would lack any semantic meaning. The computational model of unconscious processes is less worrisome mostly in that there seems to be less to "lose".

Luckily, these concerns are only relevant with respect to certain reductive claims, and my general strategy is to try to side-step them by talking like an epistemologist. Consider, for example, the epistemic norm against holding beliefs with contradictory implications. That norm is not absolute; its extent depends in part on what implications a person is capable of "seeing". Suppose I were to claim that that norm is primarily accounted for in terms of two others: one that requires holding every belief that is a consequence of one's belief—typically called a "clo-

12. Such studies are conventionally part of "neuroscience", which has become a bit of a misnomer. Someone who studies muscle fatigue at the level of nutrition and exertion would raise eyebrows by calling her work "cellular biology". But the prefix has now spread to other fields, such as "neuro-economics", which typically amounts to behavioral economics with a dusting of fMRI.

sure" requirement—and one against holding two directly contradictory beliefs (such as that it is now raining and that it is now not raining). Given this theory, the contradictions that one cannot "see" might then be explained in terms of exceptions to the closure requirement. And the latter exceptions might in turn be explained by limitations of whatever cognitive capacity *approximates* closure as best it can, given limited time and resources. This combination of norms would function together identify groups of beliefs with contradictory implications by "picking out" chains of reasoning leading back from direct contradictions, so that some further set of norms or capacities could function to revise them.

Setting aside the accuracy of this claim, it has the form of an argument that a combination of two norms can meet a third, and that the approximation of two norms by respective cognitive capacities can approximate the third. This way of talking is not far removed from my earlier description of the belief-producing machine, or for that matter from some kinds of computer engineering. It amounts to a kind of partial reduction to specific functions and responsibilities, so that one might account for one function or responsibility partly or completely in terms of another. And importantly, it leaves the question of further reduction entirely open. The questions of whether, and how efficiently in terms of time or resources, a computer could approximate closure, or even whether our capacity for closure (on the assumption we have one) does or does not depend on some non-physical "substance", is irrelevant to the analysis. Even what is conscious and what is not conscious is largely beside the point of this kind of reduction. It is *relative*, and the "lower-level" capacities are stipulated.

The extent to which I use computation as a metaphor in this thesis is compatible with this sort of partial reduction—of reduction to *parts*. The value in taking a computer (or just "machine") perspective comes from considering the contribution of smaller parts or capacities in relation to the whole. What, for example, are the pre-conditions for a certain step of epistemic reasoning, and what specifically does it accomplish? If one of those preconditions is replaced, in effect creating a new but closely related "part", what would the new part accomplish, and how might it be "used" to good effect? If it were used in the process that links ends and means, does that process take roughly the same time and effort for people to accom-

plish as the corresponding epistemic case? Do both processes seem to have about the same conscious/unconscious split? Do they tend to fail in similar ways, and at similar rates? We have some evidence to answer such questions, and could design new experiments to get more. Examining our capacities in this way allows us to consider the ways in which they might be related.

The other theory of action that has traveled farthest down this road is Cognitivism. The "part" that Cognitivists borrow from epistemology is *belief*, and the earliest "modifications" were the throwing out of one or another requirements, such as that a belief should be warranted. Kieran Setiya has more recently provided the most extensive suite of modifications and additions along these lines, by proposing a "desire-like belief" and a warrant-like requirement for non-propositional "knowledge how"[13]. Still, at the core of all Cognitivist theories lies the attitude towards a proposition that it is true[14]. The approach therefore goes against common views about direction of fit, and the ordinary view that what an agent does is up to her.

Anscombe's example of the shopper and the detective contains the seed of a different approach, one that has at its core an attitude that, like belief, is made appropriate by a kind of judgment or assessment, but with the fit appropriate to intentional action. Built around that core would not be epistemic norms and capacities, but analogous norms and capacities arranged in analogous ways. In this thesis I follow this approach and try to alter and rearrange the parts of epistemology to make the machine "run backwards", so that it can start with the representation of a goal and realize it in the world. This new machine bears the same relation to the extended intentional action function of Section 1.2 as the original does to the knowledge function. It maps a proposition to its corresponding states of affairs in virtue of a kind of warrant, which is not an input or output but an intermediate "data structure" that it produces and consumes as part of its functioning. The potential for unrealized goals or waywardness

13. See Section 3.2.2.

14. Or perhaps, more generally, an attitude towards a representation that it is accurate or "veridical".

reflects limitations of the machine's operation and of the information available[15].

1.3 The Path Ahead

The unpacking of the last section is quite abstract, and for the most part leaves open how the argument will go. So here is a summary broken down by chapter, with an emphasis on the evidence I provide for this new theory of intentional action.

In Chapter 2 I discuss the human capacity to act with appropriate confidence. Although instances of appropriate *high* confidence provide the closest comparison to knowledge, the more general phenomenon also allows for lower degrees of confidence, and reasoning about what to do depends in part on how likely it is that one will realize a potential end. This is true both when arriving at an intention, and later when one learns that circumstances have changed or one succeeds or fails at realizing incremental means. I argue that this capacity may have received relatively little attention because the most common action concepts can be analyzed without having to account for it. In Chapter 3 I outline some of the more common theories of intention and intentional action, and evaluate their prospects for accounting for this capacity, considering both how each is described by its proponents and how it might be extended.

Chapter 4 is transitional, setting up the parallel between credence and epistemic warrant on the one hand, and technical confidence and warrant on the other. Then in Chapter 5, I outline a "model" theory of credence and epistemic warrant, including probabilistic warrant for lower-credence epistemic attitudes. The primary focus is on the distinction between justification and entitlement.

I start directly accounting for appropriate confidence in Chapter 6, in which I develop a theory of *technical entitlement*. My overall strategy is to adjust the theory of perceptual entitlement of Chapter 5 to account for appropriate confidence in an intention to perform a

15. The result is not a true computational inverse—a fascinating subject that I will not discuss here. van de Snepscheut (1993, ch. 11) provides a good introduction.

basic action. High confidence is in virtue of what I call *aim*, which is the attitude towards a proposition that one is highly likely to be able to bring it about. Technical entitlement to a basic aim, together with the constitutive requirement for *non-voluntary recall* of the attitude, make its high technical confidence appropriate. The *technical justification* of a non-basic aim, the subject of Chapter 7, is the glue that connects ends and means, along with beliefs that together secure high confidence. This form of justification uses the *same* patterns of reasoning as epistemic justification, and differs only in what *attitudes* are justifying and justified. In addition to the high confidence of *aim*, I also provide a theory of lower-confidence technical attitudes on the model of lower-credence epistemic attitudes and probabilistic warrant. The evidence for the overall theory in these chapter is abductive; the adapted epistemic theory accounts for how we can intend with appropriate confidence.

In Chapter 8, I consider whether and how the theory of technical warrant accounts for many of the norms of intention and some difficult cases discussed in the action theory literature, focusing on the work of Michael Bratman. The evidence of this chapter—not all of which is positive—is also primarily abductive. However, because I do not much extend the theory of Chapters 6 and 7 in addressing these questions, their unification with the problem of appropriate confidence is also a kind of evidence. Then in Chapter 9 I offer some tentative thoughts on the implications of the theory, and more specifically the relative priority of technical rationality, for ethical theories.

Chapter 10 addresses the larger question of intentional action. Here I propose, on the model of knowledge as (epistemically) warranted true belief in the absence of certain problems, the corresponding model of high confidence intentional action as (technically) warranted true aim in the absence of certain problems. I then argue that the phenomena of Gettier problems and waywardness—the "certain problems" of the respective models—are in fact the same problem exhibited in the epistemic and technical domains of reasoning. I also argue that the appropriate model for the intentional realization of lower-confidence technical attitudes is the analogue of Sarah Moss's recent work on the relation of lower-credence attitudes to knowledge. The evidence of this chapter is primarily from the unification of intentional action and

knowledge. I start chapter 11 with a theory of our capacity to form beliefs on the basis of intentions, including the approximate content of such beliefs and how they are warranted. Then I conclude by considering how, in light of the inverse relation, warranted aims can be the basis for what has been called "practical knowledge".

One problem that I frequently touch on but defer is how we can have confident "partial plans". The theory of appropriate confidence is therefore incomplete, and my discussion of specific cases is somewhat artificial, given that we rarely plan out every aspect of an intention before acting on some part of it. However, this deferral is largely irrelevant to the larger theory of intentional action; the success of an action can be judged after all of an agent's attempts, and therefore also her plans to make attempts, are finished.

1.4 Map and Territory

There is another important but complex issue that I largely set aside in this thesis, with the exception of this brief section and some scattered footnotes. As I have said, the inverse relationship between knowledge and intentional action is approximate and therefore metaphorical. Nevertheless, this work is in part an exploration of a specific and deep relationship between these phenomena beyond epistemology's role as a useful "reference point" for a theory of action. The best evidence for that relationship is the unification of knowledge and intentional action, in virtue of warrant and the Gettier/waywardness problems that correspond to warrant. This is evidence of a relation between *territories*, regardless of whatever maps we draw to explore them.

If communicating that relation accurately and completely were my only goal, I would do so in terms of a correspondence between *problems* in each domain. For each problem on the epistemic side, I would list the potential solutions in play in the literature, and then "translate" each solution into its technical equivalent, to the extent it could be so translated. Then, assuming the problem is already recognized in the domain of action, I would do the reverse. Presenting the theory this way would therefore require discussing most of the open issues

of epistemology and action theory in some detail. I do not want to fail to write that book, which you would anyway not want to read. I am happy to settle for an incomplete account. My greatest hope for this work is for it to start a new thread in the ongoing discussion about intention and action.

However, the extent to which I take the opposite approach may worry some readers. For the most part I build a theory of intention based on the theory of epistemic warrant of Chapter 5, and then build a theory of intentional action based on the theory of knowledge as warranted true belief without Gettier problems. Some of the choices I make in Chapter 5 have inevitably been informed by the questions explored in this thesis, but I think that each can be defended on epistemic grounds alone[16]. Even so, someone who thinks I am mistaken about the nature of knowledge may wonder about the value of a theory of intentional action that makes the same mistakes.

I therefore encourage any reader who disagrees with my choices to measure these ideas not by what I have included, but by what I have left out. First, make whatever adjustments to my epistemology that you feel are needed, paying particular attention to the *scope* of those adjustments. Second, consider what would *correspond* to those adjustments in the technical theory. The key to taking this second step is thinking in terms of the underlying problem that each feature of a theory is there to address. As a simple example, suppose that you are dubious about propositional content, but accept that humans do somehow represent the world and use those representations to reason in some sense. While I talk in terms of propositions, my more general argument is that the same patterns of reasoning are used in both epistemic and technical justification. If that reasoning does not depend on propositional representation in one realm, it need not depend on it in the other. Some adjustments to the epistemic realm will translate easily, others will take some work, and there may be more than one approach. Regardless, it is in virtue of such relations, and not my own epistemic preferences, that I encourage you to judge this work.

16. That is, I do not think that my "map" of knowledge is now gerrymandered.

CHAPTER 2

Confident Action

With the path of the thesis now described, I can move on to its primary motivating problem, which is how to account for our capacity for confident action. Humans are often highly confident that they can bring about what they intend, and this confidence is generally well placed; high confidence closely correlates with subsequent success in those cases where an agent does not change her mind. Given that we often have this confidence prior to acting, it appears to be an aspect of *intention*. However, even though this capacity is common topic in action theory, the problem is at best a secondary motivation of the most discussed theories of intention, and it is not obvious how they might be extended to better address it.

This relative lack of attention does not appear to be due to disinterest or coincidence. It instead seems to trace to certain trends and preferences in the method of analytic philosophy, which favor topics that are closely linked to the right kind of concept. I also explore this problem, and the expectation that there *must* or *should be* such a conceptual link for a question to be philosophically interesting.

2.1 Acting with High Confidence

On Carolyn's last day in a remote cabin, she hopes to hike out to one of two small towns with train stations. The local terrain is not very steep, so there are many routes to either town—some offering more scenery and challenges than others. Her pack has everything that she needs to sleep at night and eat, but the cabin is in desert country and she can only carry a bit more than half a day's water. Several streams fed by snow-melt on mountains to the west are known to be reliable sources, but the springs dotting the region are less predictable. Sitting

down with her map, she decides that the potential routes to one town would be too risky, but the other offers several good options. She records a route passing a series of springs that is also close enough to a stream to offer a safe, if uncomfortable, backup source. On her way out, two springs are only damp, but because these are not consecutive, Carolyn avoids any detours and arrives in town with time for a meal and a rest.

Alex gets an unexpected call from his friend Dana, who is in the hospital after breaking her arm. She asks him to make sure that her son Kevin gets to her father's house after school. The address is about thirty miles from town and Alex has been having car trouble, but he knows that he also has many other options. If he can't establish that the car is very likely to make the trip, he can call friends, rent a car, or arrange for a taxi or ride-share. Wanting to take this problem off Dana's mind, Alex promises to get Kevin to her father's, and that he will check in later to see if she needs anything. After getting off the phone, he goes to look at the car but realizes that there's nothing he can do to get a better handle on the problem. It also turns out that no other friend is available, so this trip is starting to look expensive. Fortunately, he remembers that a coworker is out of town on business and arranges to borrow her car. He meets Kevin at school, drives him the thirty miles, and then comes back and drops off a book and earplugs at the hospital before returning the car.

The planning of Carolyn's hike and Alex's trip in these descriptions are quite different, but the focus of each is an action for which successful execution is important. In Carolyn's case, that importance comes from a basic biological need for water: if she were to stray from her planned route, or fail to detour if too many springs are dry, she could face severe discomfort or death. By promising, Alex takes on the responsibility that Kevin gets safely to his grandfather's and the additional obligation of the promise itself—even if Kevin were to make his own arrangements, it would be irresponsible of Alex not to follow through. The planning in each case is also different. Carolyn is systematic, and she only decides to set out on the hike once her whole route is established. Alex promises on the basis of very general reasoning, some of it questionable, and ends up taking a means he hadn't considered. Still, in each case, an expectation of success seems reasonable. Although both plans include riskier aspects, those

risks are countered by safe, if less desirable, backup options. If Alex had instead just assumed that his car would make the trip, our attitude towards his promise might be different; we could rightly question its appropriateness even if the car did manage to make the journey. So we can see in each of these situations a kind of virtue, in that the chance of each plan working out matches the planner's confidence that it will.

However, responsibility in such cases goes beyond one's initial planning. It extends to following through on the plan, and adjusting it, or one's confidence, in light of changing circumstances. The care that Carolyn takes in deciding on her route would not justify her continued confidence if she put little effort into following it, or knowingly changed the plan without reevaluating the availability of water. To the extent that the terrain matches her considered expectations of it, she is responsible for not needlessly straying from her path. On the other hand, a responsible plan can support high confidence without being foolproof; there is usually some chance of things not working out. Carolyn could detour to the stream and find it unaccountably dry, or a riot could break out that precludes any of Alex's means of traveling across town. When what is normally predictable fails to be, or when there is good evidence for bad information, we are less likely to hold a person responsible for not acting as they planned to. Even in such cases, however, an agent can be faulted for not adjusting her confidence in light of unexpected circumstances. Hope and tenacity can be appropriate even in the face of a desperate situation, but unwarranted confidence is not, if for no other reason than it might interfere with taking advantage of what opportunities remain.

That we can judge that someone's confidence in an intention is misplaced suggests that the relation between the two is somehow normative. But what sort of norm is involved? One possibility is that it is sociological or cultural. Of the two situations, Alex's probably sounds more familiar, while Carolyn's plan may strike many as not nearly reliable enough to be worth pursuing. There is a grain of truth in the thought that widespread confident action is a recent enough phenomenon to be attributable to culture. If the history of humanity can be fairly characterized in terms of an increase of options over time—more techniques, more ways of communicating with more people, more things available through trade—it is recent history in

particular that has seen an explosion of not just options but *reliable* options. This trend is not just a consequence of technological progress, it has also enabled it. If a complex society had only unreliable options—no advance knowledge of when a product type will be available, or whether a bus will arrive this week—further progress would be limited by the complexity of everyday planning. Working around uncertainty takes cognitive and emotional effort, as the recent controversy over "dynamic" scheduling of work hours demonstrates.

Still, it would be wrong to conclude on this basis that confident action is a recent phenomenon, and therefore more a question of sociology or economics than cognitive science or action theory. Even when few options are available, the fraction that can be undertaken with confidence have special value. Your survival depends on doing certain things at least every minute, or every few days, or every month, and although the likelihood of doing those things is not always up to you, it often is. When the well is not dry, you can stop worrying and planning about water and put your effort toward other problems. But when too much is uncertain, you may *need* to attempt more than you *can*. This importance of confidence in action to human life is reflected in how frequently we communicate it, as when one chooses between saying "Alex will meet Kevin at school" or "Alex will try to meet Kevin at school." Our ability to share projects and rely on one another often depends in part on our assessing and expressing confidence in our intentions.

Of course, that difference of expression demonstrates just as well that we sometimes act (or attempt and *fail* to act) with less confidence. Having the capacity to try in this way is also vital to our interests. You cannot always arrive at confident means to some end. You may lack skills or information, or the state of the world may preclude confidence on anyone's part. However, in such cases you may still have means to make that end more likely, and the question of whether to take less confident means often depends on *how much* more likely you can make the end. A plan with odds of one in ten might not be worth pursuing when one with odds of seven in ten would be. Appropriateness is therefore not just an issue of having high confidence or not. Mistaking one in ten odds for seven in ten could lead to wasting time and effort. And the opposite mistake could lead to rejecting the best means available.

What this capacity for confident action must "accomplish" is therefore quite complex. For each intention that one takes on or even closely considers, it must assign a corresponding degree of confidence. That degree must be generally accurate relative to both one's skills and one's knowledge (or at least beliefs) about the circumstances now and at the time of acting. And it must adjust the degree of confidence as one learns that relevant circumstances have changed, including one's success or failure in realizing means along the way, so that one can decide on new means or give up on the end. "It" is part of *us*, of course, and its "operation" consists of a combination of conscious reasoning and unconscious sub-capacities. So an explanation of the capacity will likely include a mix of conscious thought and subconscious functions.

2.2 A "Technical" Realm

The considerations I discuss in the last section overlap with two concepts widely employed in the action theory literature, but only partially.

Assessments of confidence are an aspect of the relation of ends to means, and therefore part of "practical" reasoning. But it is important that this kind of confidence is not confused with one's sense of how likely it is that an outcome will be an ethical or pleasurable thing to do. Carolyn can plan and take her hike while remaining uncertain that it is worth the trouble, and Alex can make his promise knowing he will have to skip his AA meeting, and wonder if it would have been better this time to just beg off. Given that confidence in one's reasoning about value and confidence that one can bring about one's intention can vary independently, each must reflect a different judgment.

There is also a more specific concept that captures some of Carol's and Alex's reasoning, but it is *too* specific. "Instrumental reasoning" refers to planning that takes an end as given and arrives at appropriate means. It is not always clear whether those who employ the term take it to include ethical and value judgments between different means, or if those are judged separately. But assuming they can be separated, the discarding of a potential end for lack of

sufficiently confident means falls outside of the concept, and therefore so does the *degree* of confidence. I may really want carrot cake, but settle for cereal when there is only a five percent chance the convenience store will have cake. Or I may rightly help the injured man from the water instead of swimming towards the drowning child if I judge that there is only a very slim chance I can save her.

In Section 9.1 I argue that this use of the term "instrumental" expresses (or perhaps by now, records) a mistake. While one may defer choosing a means to an end, its status as an end always depends in part on some availability of potential means, however abstract. In contrast, judgments of appropriate confidence in means do not depend on judgments of value or ethics. I can of course waste time evaluating the likely success of means too disgusting or odious to ever take, but I could also waste time repeatedly squaring a number. My *determining* that the square of 256 is 65,536 might be the result of a mistake of judgment, but the mistake is not epistemic, and does not compromise the status of *what I determined* as knowledge. Similarly, my working out routes to school through the sewers may result from an analogous mistake, but that does not affect their status as potential routes. I could, against my interests, set out to school via the sewers.

This division between types of reason and judgment extends all the way through our assessment of intentional actions. Consider the example of Sandy, who is sociopathic and indifferent but also whimsical. Most of the time, Sandy is fed through a tube, but she is occasionally struck by a whim, most recently to sink a large boat. At this whim, Sandy plans to break out of the hospital, find a sufficiently large boat, determine a way of sinking it, and take care of incidental needs like food along the way. She selects among candidate means entirely by expedience, paying no mind to whether a given route will hurt her feet or to the number of people on a given boat. Sandy's *reasoning* is entirely devoid of considerations of value and ethics, but she is still recognizably an agent, and when she sinks the boat she does so intentionally.

David Velleman might object that Sandy's action is nevertheless not "full-blooded". Following Frankfurt, he might point out that Sandy's planning is alienated from her motives in the way that an addict's can be (Velleman 1992, §§III,VIII). But the actions of an addict can

illustrate the same point. Marty steals a car stereo and delivers it to his fence to get money for his next fix. He does all this while high on heroin, and with some sense of disconnection from his actions. To the extent he his connected to them, he realizes his life is screwed up and that he shouldn't be doing any of this. Still, by now he is quite skilled at stealing car stereos. He knows how to jimmy locks and disable many alarms, and can remove the stereo using simple tools and with little or no cosmetic damage. Therefore, from one perspective we might look at Marty's behavior and ask "How intentional is this *really*?". And from another we might note his speed and precision in opening the car door, and be impressed. However compromised and degraded we judge Marty's motives to be, we can still recognize, and attribute to him, what he accomplishes. The "full-blooded" and the skilled often go together, but neither implies the other. Was his confidence appropriate given what he knew about that model of car? Was it ethical to steal the stereo? Did doing so accord with his desires? The different kinds of reasoning we would use to answer these questions reflect different kinds of reasoning on the part of Marty or any agent.

What distinguishes reasoning about ethics from reasoning from desires is not my concern here. It is rather that both of these are separate from the reasoning that picks out a potential means to an end as such, and by which we judge how much more likely that means makes the end. This latter reasoning, although still "practical", is as separate from the former as that used to warrant beliefs. A better word for this confidence and its norms, and the one I will use, is "technical", but not in the sense that "technically" can contrast with "really", and indicate that something meets the letter of a law but not its spirit. I use it instead in its sense that traces to the classical Greek distinction between *epistêmê* and *technê*, and denotes a matter of skill or craft[1]. The term "technical rationality" therefore picks out our capacity to reason about means and ends, including judging the likelihood of bringing about an end by way of some means, and the ability to arrive at whole plans for which that likelihood is high. An end could be as immediate as operating a hand pump, or more conceptually distant, such as poisoning a

1. I owe thanks to Will Reckner for introducing me to the history and appropriateness of this term, and for convincing me that my earlier instinct that the meaning of "practical" should be narrowed to this use was misguided.

household or even toppling a regime.

2.3 Technical Confidence and Action Concepts

I have argued that high technical confidence has particular value in human life. We generally do not first decide on some plan and then consider the chances of it working out. Instead, part of the value of a plan is in those chances, and some situations and social institutions call for high confidence. If this is right, then accounting our capacity to frequently arrive at intentions that support high technical confidence is a foundational problem of practical philosophy. Even so, the problem has not been a priority in action theory since the publication of *Intention*. Interestingly, this neglect does not seem to be from either lack of awareness of the issue or controversy about its significance, as judgments of confidence are frequently described and little contested. Instead, the problem seems to trace to *how* the subject tends to be raised. Consider, as an example, this quote from Davidson (1978, p. 94):

> I think X spoke correctly and accurately, but misleadingly, when he said, "I intend to go to the concert". He could have corrected the impression while still being accurate by saying, "I now intend to go to the concert, but since I may be put in jail, I may not be there". A man who says, "I intend to be there, but I may not be" does not contradict himself, he is at worst inscrutable until he says more. We should realize there is something wrong with the idea that most statements of intention are elliptical until tempered by our doubts about what we shall in fact do when we notice that there is no satisfactory general method for supplying the more accurate statement for which the original statement went proxy. And the reason is clear: there can be no finite list of things we think might prevent us from doing what we intend, or of circumstances that might cause us to stay our hand. If we are reasonably sure something will prevent us from acting, this does, perhaps, baffle intention, but if we are simply uncertain, as is often the case, intention is not necessarily dulled. We can be clear what it is we intend to do while being in the dark as to the details, and therefore the pitfalls. If this is so, being more accurate about what we intend cannot be a matter of being more accurate about what we believe we will bring off.

Davidson's concern in this passage is whether certain intentions are truly conditional just because an agent is aware of potential obstacles (he decides that they are not). Towards the end, I read him as asserting that high confidence requires awareness of "the details"—presumably

the extent to which we have planned out our actions. Along the way, and almost incidentally —his argument does not seem to depend on it—he observes that "if we are simply uncertain ... intention is not necessarily dulled". Given that virtually no intention is *certain*, the charitable interpretation of 'uncertain" here is "somewhat in doubt". In Davidson's view, therefore, a lower confidence plan is still an intention.

This passage is one instance of a common pattern: a frank discussion of what I am calling technical confidence, noting or implying the possibility of high confidence, in the service of arguing that the explication of some concept, in this case *intention*, does *not* require explicating or accounting for high confidence.

Working out the meaning of concepts is at least part of the task of the analytic philosopher, and Davidson's observation is entirely appropriate given his view of the evidence. The pattern, however, suggests a more specific principle, which is that the best context for discussing some phenomenon is in the explanation of a concept that calls for it. One would therefore most expect to see high technical confidence discussed in working through some concept that "cuts at the joint" between high and lower confidence, expressed with a common word like "intention" or "try" rather than a made-up phrase like "high technical confidence".

It is this expectation that seems to be the root of the relative lack of attention to the problem of technical confidence in action theory. There are a few concepts that, at first glance, seem to be good candidates for providing a good context for the discussion. The consensus in each case is that that appearance is misleading. While not everyone agrees with the consensus, subsequent debate has generally focused on the conceptual question rather than attempts to account for the capacity.

2.3.1 *Do* versus *Try*

In an earlier section I said that the "importance of confidence in action to human life is reflected in how frequently we communicate it, as when one chooses between saying 'Alex will meet Kevin at school' and 'Alex will try to meet Kevin at school.'" I said "communicate" because

it is generally believed that neither of these phrases is *about* Alex's confidence. *Try* might seem to be a solid candidate for a concept that includes lower technical confidence while excluding high confidence. But the consensus view is that you also try when you are confident. There is some straightforward linguistic evidence for this theory: Suppose that Alex is indeed appropriately confident and says "I will meet Kevin at school". If there is a freak occurrence like a riot and he fails to meet Kevin, he still could rightly say "I tried". Both Hornsby (1980, §III.2) and O'Shaughnessy (1973) offer frequently cited arguments that *try* applies across degrees confidence[2].

There are similar arguments about *do* and the unqualified use of action verbs. "Alex will meet Kevin at school" is just a statement about a certain future event that will happen or fail to happen, and not even about Alex's attitudes, let alone his confidence. And anyway, when Alex tries to win at cards (with lower confidence) and succeeds, he won at cards. While Anscombe (1957, §47) notes that some verbs imply intentionality, none seems to imply high confidence. So, although we most often communicate our degree of confidence by including or omitting "try", what we say does not entail that degree, and both *do* and *try* can be explicated without addressing high confidence.

This "arrangement" of communicating confidence should be familiar, because we most often communicate credence in a similar way. When I say "the cat is on the mat", I state a fact that is not in any way *about* an attitude or my credence. And the easiest way to express some doubt is to say something like "I think the cat is on the mat." But surely "I" *think* in both cases. So, in virtue of my choice of words, I communicate my degree of credence without saying anything about it.

2.3.2 The Progressive Tense

There is one wrinkle in this view of the meaning of *do* and other action verbs, which is that their meaning shifts slightly when cast in the progressive. I tell you "I am going to the store"

2. The view is, naturally, not universal. S. Schroeder (2001) offers one counter-argument.

as I head out the door. Later, I am hit by car and taken by ambulance to the hospital, where I realize I won't be going to the store anytime soon. So I certainly did not go to the store (this time), and yet it seems fine to say that I was going to the store when I was hit by a car. Similarly, I could lay glasses out on my table and then find that my flatware was stolen. I might then be said to have made this discovery while "setting the table", even though the table was not set and could not have been set.

This aspect of progressive tenses is a subject of Parsons (1989, §§4–5), in which Terry Parsons discusses some proposals to account for it, including his own. David Dowty, he notes, argues that I would have gone to the store in the relevant "inertia world", which is a "possible world" similar to the actual one at times before I am hit by the car, and "after this time develops in ways most compatible with the past course of events." Parsons observes of this theory that the actual world seems like a good candidate inertia world, and therefore the idea needs some sharpening. I would add that any conception based on a specific *time* would have trouble accounting for epistemic problems like my missing flatware, or my heading to the store only to discover it was washed into the ocean months ago.

Parsons' own view starts from the observation: "an event often has both a development portion and a culmination. For example, if Mary builds a bookcase then there is a period of time during which the building is going on—the development portion—and then (if she finishes) a time at which the bookcase finally gets built, the time of culmination." (p. 219). If this is right, then perhaps the progressive tense refers to the development portion of an event, whether or not it culminates. One difficulty of this approach are "objects" that never exist but are *referred to*, such as a bookcase Mary "is building" when she discovers her sawmill has no power and she can't make her tree into boards. Parsons offers his own view of how to address such cases, including the problem of when "the development portion" really starts[3].

The progressive may be relevant to the issue of high technical confidence because of how it muddies the contrast between *doing* and *trying to do*, given that the former is apparently not

3. He also discusses Michael Bennett's analysis, which I take to be close to his own with respect to the issues relevant here.

the occurrence (or "culmination") of some event. Suppose I say "I'm trying to go the store" as I head out the door on crutches. On this trip I also get hit by a car and am taken to the hospital. According to the conventional view, I *try* in either case. How, then, does "I am going to the store" differ in truth conditions from "I am trying to go to the store", if I can do either without winding up there? *Do* the truth conditions differ? On Parsons' view, it is not obvious to see how they would, unless the basis of difference is somehow located early in the "development portion". Dowty may be in a better position to forge a difference, assuming that the concept of "inertia worlds" can be productively refined: *trying to do* might be true in inertia worlds where *doing* is false.

Still, these theories of the difference in meaning are speculative and need further refinement, leaving the field open for other interpretations. Another possibility is that what is normally merely communicated by "do" is part of the meaning of *doing*—that the truth conditions of *doing* depend on how much the agent is "on top of" her situation. Suppose it is true that I can be setting the table when I find that my flatware was stolen weeks ago. Would that claim become dubious if I were instead aware that a friend often borrows my flatware without asking? If it is more dubious, then the measure of my being "on top" of the setting may just be my high technical confidence and its appropriateness in context—that is, my being confident and my having arrived at that confidence properly.

However, this aspect of the progressive has been of more interest in philosophy of language than action theory. When it does come up in the latter, it is often as a means of loosening a requirement. Knowledge of what you *are doing*, for example, might be easier to achieve than knowledge of what you *do*[4].

2.3.3 *Intention*

Another candidate for a concept that implies high technical confidence is *intend*/*intention* (assuming these are the same). Michael Bratman argues that the primary cognitive role of inten-

4. See, for example, Paul (2009a, p. 16).

tion is to coordinate future plans. To play that role, intentions much be consistent with one's beliefs and also with each other. I should not plan to go to the store if I know it has been washed away, or if I also plan to set the table at the same time. Either of these would be a *rational* mistake.

Bratman takes these requirements of consistency to apply to intentions generally, and on that basis argues against the "Simple View" that everything one does intentionally is preceded by an intention to do it. I discuss his famous video-game argument for that conclusion in Section 8.3.2. Here it suffices to note that there is a sense in which one can rationally have incompatible goals. I can, for example, put effort towards making a chocolate cake and also a yellow cake knowing that I only have enough frosting to finish one of them. Which cake I wind up with might be determined by a mistake during baking or, if both go well, my eventual decision. However, although my "cake situation" is disjunctive, it does not seem that my efforts to make the cakes need to reflect that. I am not required to bake-a-chocolate-cake-or-a-yellow-cake; I can just proceed with the chocolate cake as if the yellow cake were not part of the picture, and vice versa.

In his description of the difference between rational and irrational conflicts between goals, Bratman makes two points that suggest a possible link to high confidence. One is that he associates the stronger commitment of *intending to* do something to the belief that *one will* do that thing (Bratman 1984, p. 384). The second is associating the weaker degree of commitment with *trying* (p. 387): "My response is to reject the contention that I must intend to hit each target in order to best pursue the reward. What I need to do is to *try* to hit each target. But this does not mean that I must *intend* to hit each target. Perhaps I must intend to *something*—to try to hit each target, for example." Associating Bratman's view of *intention* and high confidence also makes intuitive sense. High technical confidence is inappropriate in light of incompatible beliefs or plans, while lower degrees of confidence allow for conflicts. To take a simple case, suppose that my baking skills only allow me a forty percent chance of making a given cake. That attitude would seem to leave my two-attempt strategy rational in the way that Bratman envisions.

I opened this section with an argument from Davidson (made prior to Bratman's) for the opposing view. That is certainly not the only argument along those lines. Another strategy, too complex to discuss here, starts from Davidson's view of identity of action, together with a claim for appropriate confidence in "basic action", which is the topic of Chapter 6. But the force behind these arguments—the source of the expectation for why they *should be* true—is an expectation against equivocation. The "Simple View", after all, is an analysis of the link between *intention* and *intentional*. When I try to go to the store and wind up there, I (generally) go to the store intentionally. To say that one can not be intending to go to the store when one tries to go to the store is to say that some of what is intentional is not preceded by a corresponding intention. Such equivocation is often considered philosophically fishy. Indeed, Anscombe devotes the very first paragraph of *Intention* to flagging as implausible any claim that *intention* and *intentional* are equivocal[5] (Anscombe 1957, §1).

There are also suggestions of a link between the *low* end of technical confidence and the concept *intentional*. Suppose I play the lottery and win. In that case I may have tried to win the lottery, but it is at least strange and plausibly wrong if I claim to have *won* it intentionally. A natural way to put the problem is that my winning is too much a matter of luck, but what counts as "too much luck"? One possible answer is that one's degree of appropriate technical confidence is too low—that for an action to be intentional one must have a minimal handle on the process of bringing it about. But there have been other suggestions, including that for an action to be intentional one must apply a certain degree of skill in bringing it about[6].

5. Other philosophers accept that the evidence favors some form of equivocation. Velleman (2007, pp. 112–3) argues that 'intention' can be used either in Bratman's sense or a different sense that describes any goal, and says, "In ruling these terms ambiguous, I am giving up, once and for all, on the project of finding a unified analysis for the expressions in which they appear." Audi (1973, p. 388) says, "There are some people who will not say they intend to do something, or bring about some state of affairs by doing something, unless they are quite sure that they will do, or bring about, whatever it is. Others use 'intend' less stringently."

6. See Audi (1973, pp. 387–8) and also Section 10.2.5.

2.4 Concepts and Philosophical Method

Even if ordinary action concepts do not express technical confidence, it is easy to talk about—one can just say "I'm confident I can do this" or "I'm beginning to doubt whether I can pull this off". Given its importance to intentional action, therefore, it may seem that nothing further stands in the way of my developing a theory of theory of technical confidence. The conceptual approach runs deep, however, and some of the methods of this thesis diverge from it. So I end this chapter with a few words about that conflict, working back from what I take to be the problem.

In Section 1.2.1 I note the link Anscombe draws between failures of performance and knowledge of one's situation. She also famously claims that we have "knowledge without observation" of our intentional actions. Setting aside the difficulties of "without observation", *knowledge* carries at least an implication of high confidence. And throughout the text, most of the examples of intention she describes are either the sort of thing one often does with confidence, or could be. But here is one passage that includes an interesting exception (Anscombe 1957, §23):

> Now if all this holds, what are we to say about all these many descriptions of an intentional action? ... E.g. 'Why are you moving your arm up and down?' 'To operate the pump', and he is operating the pump. 'Why are you pumping?' 'To replenish the water-supply' and he is replenishing the water-supply; 'Why are you replenishing the water-supply?' 'To poison the inhabitants' and he is poisoning the inhabitants, for they are getting poisoned. And here comes the break; for though in the case we have described there is probably a further answer, other than 'just for fun', all the same this further description (e.g. to save the Jews, to put in the good men, to get the Kingdom of Heaven on earth) is not such that we can now say: he is saving the Jews, he is getting the Kingdom of Heaven, he is putting in the good ones.

The substance of this "break" does not appear to be temporal—in her example we are given no reason to think the inhabitants are ingesting poison as the pump arm is moving. It rather seems to relate to what we can judge at the time: whether the Jews will be saved or the good ones will be put in is, compared with the pumping and the poisoning, more doubtful.

This is hardly proof that Anscombe was mostly thinking of high-confidence cases when writing *Intention*; I personally doubt that the issue was at the front of her mind. However, it does raise questions about the status of this passage from Davidson (1963, pp. 91–2) as the "standard refutation" of her knowledge claim: "It is a mistake to suppose that if an agent is doing something intentionally, he must know that he is doing it. ... in writing heavily on this page I may be intending to produce ten legible carbon copies. I do not know, or believe with any confidence, that I am succeeding. But if I am producing ten legible copies, I am certainly doing it intentionally."

My worry is not with Davidson's wording[7], it is that whatever the wording, his argument is still ultimately terminological. Davidson is right that when one's lower-confidence attempt succeeds in the right way, its realization counts as something one does intentionally. However, nothing prevents our charitably reinterpreting *her* claim as applying only in cases of high-confidence. We can say: "Yes, Anscombe was wrong to put things as she did, but many of her ideas can be carried forward by differentiating between high and lower confidence intentions." On the epistemic side, that restriction is standard—our also having lower credence attitudes is no threat to the phenomenon of belief-based knowledge. In fact, epistemology has traditionally focused so much on high credence attitudes that comprehensive accounts of probabilistic reasoning are a relatively recent development.

The work in this thesis is in part an attempt to carry forward some ideas from *Intention*, and I trace my interest in technical confidence directly to Anscombe's work. But given its wide influence, I cannot help but feel I should be joining a long and lively discussion of the subject rather than fretting over the space I hope to carve out for one. It is not as if the conversation never started for lack of an ideal concept, because the subject is in plain sight in the dialectic between Anscombe and Davidson. Nevertheless, philosophical interest in the general area of technical confidence has been mostly relegated to Cognitivism, which as I note in Section 3.2

7. A counter-argument based on his use of the progressive is not out of the question, but that approach would not obviously help Anscombe's case, given that she first introduces the idea this way: "Now the class of things known without observation is of general interest to our enquiry because the class of intentional actions is a sub-class of it." (Anscombe 1957, §8).

is distinguished in part by having a specific conceptual rationale.

The expectation of conceptual analysis seems to continually guide the allocation of philosophical resources towards one kind of question and away from others. I have seen the everyday version of this force in feedback on my own work. It most often takes an unambiguously helpful form. For example, in earlier drafts of what is now Section 2.1 I said that appropriate technical confidence is part of what makes a promise appropriate. Readers responded with pointers to Marušić (2013), which discusses some promises that one cannot be appropriately confident in, such as that one will quit smoking or continue to love someone. I understood the message—the relation I posited was faulty if at least *some* promises do not require confidence. At the same time, promises of that kind are not really relevant to the larger argument. So I addressed the issue by substituting examples for the generalization, leaving each reader to generalize (or not) as she sees fit.

Feedback like this is more ambiguously helpful when it continues after the relevant corrections are made. And it does continue. When I send out a draft that describes a *particular* promise, and claim that part of what makes it appropriate is a certain kind of confidence, inevitably some reader will still suggest that I have a look at Marušić (2013). If I posit any communicative link between self-described trying and confidence, without making any claim about the meaning of "try", a reader will express "surprise" that I did not cite Hornsby's analysis. Some of this advice may trace to personal mental summaries, which can muddy the examples and concepts in a text. But however they arise, these nudges are how the expectation of conceptual analysis is transmitted and enforced by proxy. The implication is: "Your other readers will take you to be doing conceptual analysis, so I am reading your work as if you are. On that reading, here are some holes in your analysis."

I think it is fair to ask whether philosophers should expect a paper to feature all of this conceptual analysis, to the point of distilling it on behalf of an author. But I do not care much about that question. I still frequently work in the conceptual mode, especially when approaching a new subject. If there is Good Advice for Philosophers to be had here, I leave it to others. My problem is a personal one: When it comes to these action concepts, I simply

no longer have a position on these questions. I achieved this ambivalence honestly, in trying to carry out the expected analysis. In earlier stages of this work, I had views as to whether one can rationally have incompatible intentions, exactly what "try" means, and so forth. Like Davidson, I claimed that certain phrases "sounded misleading" without being "wrong". In trying to support these views, however, I largely argued myself out of them. The evidence from usage alone seems at best ambiguous. Some arguments against equivocation in one concept (or, I suppose, concepts) wind up imposing it on another, and therefore just moves the problem around. I eventually admitted to myself that I was picking from this evidence with my preferred theory in mind, which amounts to confirmation bias.

For what it is worth, I can offer two possible explanations for why the words we use to describe what we do might have messier meanings than, for example, epistemic terms. One just follows Wittgenstein's metaphorical observation that the oldest parts of a city are the most irregular (Wittgenstein 2010, §18). Action does not precede representation—it seems most likely that these arose together and are necessarily intertwined. However, action certainly precedes language, and communicating intention must have been among the very first uses of language. Communicating beliefs *as* beliefs may have been another, but the mode of stating facts could have sufficed for some time. In that light, is it surprising that a change of tense could have something like the effect of equivocation?

The other explanation comes from the potential for interaction between concept and theory. Philosophers tend to assume that concepts are prior to theory—that any discrepancy between the two must be resolved by revising the theory. But over long enough periods, there may also be some pressure in the other direction. Is it striking, for example, how little we now worry over non-factive uses of "know". While I think I agree with the common view that such uses are just incorrect, I also wonder whether several millennia of a theory marking them as wrong is part of what makes it easy to write them off now[8]. Theories of action,

8. This is distinct from a concern that philosophers may be choosing linguistic evidence based on theory. The thought is instead that the theory was originally *roughly* right, and the recognition of its rightness over the years caused the meaning of the concept to gradually drift further towards it. This process would make the theory weakly "performative" in the sense of MacKenzie (2008, ch. 1).

in contrast, have been more partial and less influential, and may have left more conceptual messiness intact.

Either way, my purpose in detouring through these weeds is not to argue about what cannot be done, but what can. The best lesson of my acquired ambivalence is how little difference it can make. Consider again Bratman's two-part argument that intentions should be consistent. The first part defends a certain kind of practical reasoning that includes a norm against forming incompatible plans. The second part argues that this kind of reasoning is linked to the concept *intention*. The expectation of conceptual analysis tends to focus effort on confirming or denying the second part. However, if we do sometimes reason as Bratman describes, then an account of how that reasoning works would have value independent of the conceptual question. And if we also sometimes reason in a different way that allows plans to conflict somewhat, an account of *that* reasoning would also have value. An integrated account—one that allows reasoning of the first kind to depend in part on reasoning of the second kind—would be more valuable still.

My sense of the current state of action theory is that there is less disagreement about how it "works" than about how to draw lines *through* how it works, to best align with our concepts. I do not mean to imply that there is no disagreement about the former, but only that there is plenty of room for developing a theory that is less closely tied to conceptual questions. My method in this thesis is one such approach. My main interest is in similarities and distinctions among capacities. Conceptual analysis has played a vital role in bringing these similarities and distinctions to light, and I do a bit of it. But for the most part I stipulate names for the various features of our capacities and then move on to the problem accounting for them.

Theories of this sort can contribute to further conceptual refinement. However, that is not my motivation, and I am no longer confident that those questions have much further philosophical value *in action theory*. If the correct meaning of *intention* were pinned down, it could provide evidence for one theory over another. But we first need theories that are robust enough to draw comparisons between. The chapters to come may succeed or fail by that standard. What I hope to have made clear and defended is that *it is a standard*.

CHAPTER 3

Technical Confidence and Established Theories

The primary claims of the last chapter are that appropriate technical confidence is important to intentional action, and that the subject has received less theoretical attention than it should have, given that importance. In this chapter I examine some of the attention it *has* received. Most theories of intention include some provision for differing degrees of confidence, but the extent to which they address its appropriateness varies. Because the problem itself is marginalized, it can be unclear whether appropriateness poses a real problem for a given theory, or if the theorist was just less concerned about that question. I therefore evaluate each theory both as it is presented, and also how it might be altered or extended to better account for appropriate confidence, in light of three questions:

1. How is technical confidence "encoded" in, or otherwise associated with, an intention?
2. What makes that degree of confidence appropriate at the time the intention is formed?
3. How is the degree, or the intention itself, revised in light of changing circumstances, including failing to bring about an intended means?

This chapter is therefore, at root, a survey, which raises the specter of due diligence. If the only question were whether each theory is deficient, however, I would have put this material in an appendix. The value of this analysis is the picking out of what the theories do address, and specifically how they address it. I therefore focus on *features* of theories and the limitations of those features—on what works and what doesn't. Some of these features reappear, directly or slightly altered, in later chapters. The limitations of existing theories also illustrate the contours of the problem of technical confidence, and make it specific.

This chapter includes some epistemic terminology that I go on do discuss in Chapter 5, so it may help to skim parts of that chapter as you read. I also use an idiom from the action theory literature: In the way that "n" stands in for *some* integer in mathematical contexts, "φ" (the Greek letter *phi*) is often used to stand for some action—most often a (relative) means to some end. When it occurs as part of a mental state, it can be thought of as a bit of mental content, perhaps part of a proposition. The grammatical context of the symbol should not be considered to restrain its application. For example, "φing" can mean "making it the case that φ", where "φ" stands for some propositional goal. I use "φ′" and "φ′′" for additional, distinct actions as needed. I also refer to "bringing about A", where "A" is some propositional goal, typically a (relative) end.

3.1 Belief/Desire Pairs

One of the most influential ideas in twentieth century action theory is that an intention consists of a desire paired with a related belief. It is generally credited to "Actions, Reasons, and Causes", in which Donald Davidson (just about) gives the example of a desire to stop the main from backing, and the belief that easing the jib will stop the main from backing (Davidson 1963, p. 6). Most authors following Davidson use examples that match this template, in which a desire for some end A is paired with a belief that φing (a means) will bring about A. Beyond this basic pattern, however, it is difficult to generalize about what is really a *family* of belief/desire pair theories. Davidson's original gloss is (p. 5):

> R is a primary reason why an agent performed the action A under the description d only if R consists of a pro attitude of the agent towards the action with a certain property, and a belief of the agent that A, under the description d, has that property.

However, this phrasing is laden with his event metaphysics and theory of the identity of actions. The former has turned out to be much less popular than the pair theories, while the latter is still widely held but controversial. To that template, Davidson also adds the requirement that the intention (or "primary reason") be the *cause* of A. This separable claim is now

known as "Causalism" and is if anything *more* influential and widely-held than any particular reduction of intention to other mental states.

Davidson's initial version of the theory raises two questions: Doesn't a person often have desires they never decide to act on? And isn't someone often aware of more than one potential means of bringing about some end? I can want to go to Vienna *in some sense*, while never really having a trip to Vienna as my goal. Or I can plan to go there by plane while also knowing I could travel by train. At first glance, at least, the belief/desire template seems to pick out too many states as intentions. Davidson's thought at the time was that Causalism would address this concern: Trace the causal chains back from some event, and you may reach a paired belief and desire. If you do, that pair is likely an intention[1]. The specifics of how such pairs come to initiate the causal chain might then be set aside, or at least contained in the problem of what is not "deviant".

The best evidence that his version of the pair theory took this simple form comes from the reason he later gave for abandoning it, which is that it does not account for what he called "pure intending", as when one intends to φ and then changes one's mind before acting (Davidson 1978, p. 88). The combination of the pair template and causal selection therefore picks out too *few* states as intentions. Other theorists have responded to this problem by modifying the theory. The causal chain is typically retained, but now as a requirement on intentional action, and other criteria are given to select some matching pairs of belief and desire over others. Some of these theories also tinker with the template, as with this (slightly rewritten) theory from Audi (1973, p. 395), which is one of the more widely discussed variants[2]:

> x intends to bring about A by φing if and only if (1) x believes that he will (or that he probably will) bring about A by φing and (2) x wants, and has not temporarily forgotten that he wants, to bring about A by φing, and (3) either x has no equally strong or stronger incompatible want (or set of incompatible wants whose combined strength is at least as great), or, if x does have such a want or set of wants, he has temporarily forgotten that he wants the object(s) in question,

1. Although he realized this procedure is not foolproof, due to the problem of "deviant" or "wayward" causal chains, which are a subject of Chapter 10.

2. Sinhababu (2013) includes a more recent specification.

> or does not believe he wants the object(s), or has temporarily forgotten his belief that he cannot both realize the object(s) and bring about *A* by φing[3].

Audi's theory "selects" those desires that are strongest when compared to our other desires taken as a whole. It addresses the problem of belief selection by "repeating" the belief in the desire—by making a preferred means part of what the agent (most) wants. This approach does put certain constraints on considerations of technical confidence, and I discuss these below. For the most part, however, the criteria by which matching pairs are identified as intentions or not are not very relevant to the confidence problem. For present purposes, we can think of the members of this family of theories as *somehow* picking out matched pairs, and consider how one could have appropriate confidence in an intention of that form.

The historical reasons for the influence of the pair theory are not my subject here, but some attractive properties are plainly visible in its structure. First, to intend to bring about *A* is to take some sort of positive attitude towards *A*. Whether or not that attitude is precisely a *desire* for *A*, the pair theory posits that an intention includes a separable pro-attitude towards *A*, presumably governed by pro-attitude-appropriate norms. Second, as I have argued using the terminology of technical confidence, an intention to bring about *A* also depends on having some handle on how to contribute to *A*'s coming about. The paired belief about one's contribution to realizing *A*, which is also separable and presumably governed by epistemic norms, responds to that intuition. One suggestion or hope of this family of theories is therefore that two already familiar attitudes might together play a third role of constituting an intention.

Another advantage of the separable belief is how it addresses the first two questions of technical confidence. Consider first the potential "encoding" of *high* confidence. Assuming an available means of easing the jib, a belief that easing the jib will stop the main from backing provides a high-confidence path to stopping the main from backing. That confidence is more specifically in virtue of the high credence that is characteristic of belief. Taking this approach to technical confidence as a given, there are two general options for lower-confidence intentions, corresponding to sides of a debate in epistemology. One is Audi's choice, which is to

3. Audi gives a different specification for "simple" intentions, which are those for which there are no further means.

express lower confidence in the content of a (high credence) belief ("or that he probably will"). Another is to keep the content fixed and take a lower-credence attitude towards it[4].

Similarly, the question of appropriate confidence in arriving at the intention can be answered in terms of what makes the credence of any belief appropriate: that it is epistemically warranted. This is a very general answer, in that there are different kinds of warrant, and the warrant for beliefs like "easing the jib will stop the main from backing" may pose particular difficulties. However, at least at a high, structural level, it is sound. And anyway, there are familiar stories for how beliefs like this can be warranted. You might be an expert on the theory of sailing, and can deduce this principle from more general knowledge. You might have personally seen this particular approach succeed on hundreds of occasions. Or you might have learned the principle from a trusted source. The advantage of basing intention in part on belief is that with belief come epistemic norms and the mechanisms that account for how we generally meet those norms.

It is with respect to the third question that belief/desire pair theories start to falter in addressing technical confidence. Suppose that my intention starts as the combination of a desire for A and a belief that φing will bring about A. The belief is presumably general—that *any* φing (under relevant circumstances) will bring about A. My eventual action, however, will be specific; I will need to φ and do so intentionally. Even setting aside familiar concerns about infinite regress and assuming I can just φ, I still need to intend to φ in order to do so intentionally. The question of ongoing confidence is therefore the question of what links my intention that A to the success or failure of my intention to φ.

One way to conceive of this problem is as another gap in the Causalist picture. A "successful" intention to bring about A kicks off a causal chain that eventually realizes A. One step along that chain will presumably be an intention to φ. If all one cares about is the existence of this chain, the detail of what states or events lie along it might be left open. But if one cares about the ongoing maintenance of technical confidence, the specific relations of some

4. I may be over-interpreting Audi by attributing one of these two views to him based on the quoted passage. This encoding choice has little to do with the concerns of his paper.

of the "downstream" states to the initial intention become important. Appropriate confidence does not just depend on the general belief that easing the jib will stop the main from backing, it also depends on the success or failure of easing the jib. And if all technical confidence is to be encoded as belief, it seems inescapable that confidence in an end eventually needs to be epistemically warranted in part by an intention to carry out a means, and therefore by a desire. However, this relation, of a belief depending in part on a desire, breaks the clean separation of pro-attitude and epistemic attitude that is supposed to characterize this type of theory in the first place.

Audi's variant theory suggests a way out of this problem, which building the end and means into a unified attitude. If what I want is to *A*-by-φing, I might somehow maintain a sense of confidence in my ongoing desire to *A*-by-φing. However, even if this everything-at-once approach could be made to work for "single-step" intentions, it is not clear what relevance it has to intentions that take multiple steps to realize. I might intend to bake-a-cake-by-measuring-flour-and-then-measuring-butter-and-so-on, but when it comes to taking actual steps I'll first measure the flour. I need to assess confidence in terms of the success of individual steps, and in light of changing circumstances. Belief/desire pair theories do not account for how we do this, and there is no obvious way to do so without breaking the encapsulation that defines them.

3.2 Cognitivism

"Cognitivism", as the term is used in the context of action theory, also denotes a diverse family of theories that fit a pattern, although there is no consensus on exactly what that pattern is. The quick but not entirely accurate slogan is that intentions are a variety of belief, but not every Cognitivist theory claims that the state consists *only* of belief. The more ecumenical version is that intentions "involve" beliefs, but under that criterion belief/desire pair theories would qualify, and those are not generally considered to be Cognitivist. The now preferred generalization—I do not know its origin—is that a Cognitivist theory accounts for certain rational

requirements on intention in terms of rational requirements on belief. Ross (2009, p. 243), for example, proposes requirements of *consistency* and *means-end coherence*[5]. However, although these specific requirements are a focus in the recent literature, they were not obviously so early on. The initial motivation was instead just the claim that when one intends to φ, one must believe that one will, or will *probably*, φ—the simplest explanation for a necessary connection being an identity (Harman 1976, pp. 432–5). This connection is sometimes put forward as the accurate part of Anscombe's knowledge claim[6].

There are too many Cognitivist theories to discuss the whole family in detail. However, just as with pair theories, many of the distinctions and debates don't much interact with the problem of technical confidence. I will therefore consider two variants. The first is a generalization of what might be called "early" or "standard" Cognitivism, consistent with the features cited in Ross (2009)[7]. The second is Kieran Setiya's variant, which differs from the standard view in ways relevant to the three questions.

3.2.1 Traditional Cognitivism

On its own, the identity claim suggests that an intention to φ might just be a belief that one will φ. I can, however, either intend to fall down the stairs or believe that I will fall down the stairs, and these seem like different attitudes with different implications. If intentions are beliefs, they need some content to distinguish them from predictions based on evidence[8]. The standard suggestion, which is common to a surprising number of distinct Cognitivist views, is that the additional content is or includes some kind of self-reference. Gilbert Harman proposes something like "I will φ because I have this very belief", and David Velleman's offers "I will φ

5. See Sections 8.2 and 8.3.

6. See, for example, Setiya (2008, §1).

7. My interest here is only the interaction of Cognitivism with appropriate technical confidence, and not in the advantages and problems of such theories more generally. However, the Cognitivist approach does face significant challenges, and Ross's paper provides excellent explanations of a number of them.

8. Velleman (2007, pp. 98–9) argues that an intention is a kind of "reflective" prediction, but still recognizes the "commonsense distinction" between intentions and ordinary predictions.

herewith" (among other forms) (Ross 2009, pp. 254–5). The benefits and drawbacks of different forms of self-reference have been a locus of much debate, but the issues do not much interact with technical questions.

Even on the simple picture of a somehow-self-referential belief, it is clear how high technical confidence would be encoded. The answer is similar to that for pair theories: it is encoded in virtue of the belief, or more specifically the credence in the belief. With respect to lower confidence intentions, Velleman follows Harman in preferring to adjust the content rather than the credence (Velleman 2007, pp. 114–7). Such an intention would therefore follow the pattern "I will contribute to bringing about A (or "try" to) in virtue of having *this belief*." Even if the intention is most directly related to the activity that makes A more likely, Velleman argues that A is still the "goal" of the intention (p. 116), which is enough to make the agent's bringing about of A intentional.

Self-reference does more than distinguish intentions from predictions. It also puts the relation between the belief and the eventual realization of the goal into truth conditions of the former[9]. Such a belief gestures toward whatever future activity on the part of the agent will make it true. Indeed, in Velleman's picture, the motivation to follow through on one's intentions amounts to the motivation to have accurate beliefs (ch. 1), although this is an intermediate and formal motivation. One arrives at an intention to φ because one has practical reasons to φ. Only then does one try to make one's intentions true[10].

Another function of the self-reference is to partly account for the most controversial feature of Cognitivist intention-beliefs, which is that they are not epistemically warranted in any normal sense. Actions are not inevitable; we do not deduce our plans from our motivations and what we know of our situation—otherwise intentions *would* be predictions (pp. 96–9). They are instead voluntary, so if intentions are beliefs, what makes that attitude appropriate is our

9. Along the same lines, Causalism can be made explicit in a Cognitivist framework if the posited belief is something like "I will φ in virtue of this belief causing me to do so."

10. On this view, the second way of encoding lower confidence is probably unworkable. When credence is also one's *motivation*, a lower-credence attitude towards φing is compatible with not even attempting to, which goes against the cognitive role of intending.

general tendency to *make* them true. Therefore, whatever it is that makes initial confidence in an intention appropriate cannot rest on standard epistemic norms, as it did in the case of pair theories. This difference introduces two distinct complications. The first is formal: A common objection to the possibility of a self-referential belief is that one cannot rationally adopt it (Ross 2009, §2). If the credence in the belief ultimately rests on an irrational foundation, so does the technical confidence. I share some of these doubts, but have nothing to add to them here.

The other, more pressing complication is the matter of what other norms govern the adoption of such a belief in lieu of a warrant requirement. Some of these norms presumably relate to motivation, but they cannot all do so. I may very much want to travel to Mars, but it is very doubtful that I could rationally adopt the belief that I will travel to Mars in virtue of having this very belief. Perhaps I can rationally *try* to travel to Mars (and perhaps not), but some factor must at least rule out high confidence on my part. Cognitivists most often offer two requirements in response to this worry. The first is the distinct epistemic requirement against adopting inconsistent beliefs. If I believe that A can't come about, or that I can't φ, I should not intend that I will. The second requirement is that I know how to φ, or perhaps, if there is time, that I know that I can know how to φ. To this point Velleman (2007, p. 125) says, "To be sure, an agent isn't entitled to intend to do something unless he knows how.", which he footnotes with "Of course, he needn't know how to do something in order to intend on doing it *later*, since he can intend on acquiring the requisite know-how in the interim."

Even when restricting the problem to judgments of high confidence, however, the combined impact of these two requirements is vague in comparison to a warrant requirement. The difficulties lie in the *interaction* between "knowing how" and knowledge of one's circumstances. I may be a good driver in good weather, but dodgy in the snow. Snow does not make it *impossible* for me to get to the store, even if I plan to drive, but I should not be highly confident about getting there. It might be objected that my proposed intention fails to pick out the relevant skill, which is driving-in-the-snow to the store. If I thought of the goal that way, I would, or should, lower my confidence. However, just because I can think of the problem in this way

does not explain why I must, especially if the only limits on a valid plan are individual skills and a need to avoid inconsistency. If simply driving to the store is to be *inconsistent*, I must take myself as having two distinct skills of driving-in-the-snow and driving-not-in-the-snow. (One can't drive-not-in-the-snow in the snow, after all.) Or rather as having distinct classes of skill, as every other differentiating condition, such as whether it is light or dark outside, must presumably cleave my abilities in the same way.

Here is another way of expressing what I take to be the same problem: Judgments of technical confidence can be rightly called "judgments" in part because we often have reasons for them. If I promise to get you baby formula, grab my keys, and start out into the snow, you might ask me whether I am in a position to make that promise. Do I plan to drive, and do I remember all the times I have gotten stuck and had to wait or abandon my car? Any judgment of confidence, but particularly those of high confidence, should be responsive to (epistemically warranted) beliefs about the situation, preferred means and potential alternatives, skills, and how all of these interact. When you argue that I should put more weight on my past experience, you are giving me a reason my confidence should be lower. The problem of appropriate confidence can therefore be seen as the problem of providing the requirements on that kind of reasoning.

In the realm of knowledge, the negative requirement against contradiction does factor into the analogous picture, but most of the account of how evidence "interacts" is in virtue of the positive requirement for warrant. By removing the latter, Cognitivism fails to provide adequate norms for high technical confidence. The picture is even worse with respect to gradations of lower confidence, which can also be relevant to the question of what intentions to adopt. An eighty percent chance of getting formula in the next two hours may be worth taking, when a five percent chance would not be. The general sense in which I can be said to "know how to get baby formula" provides little help in deciding what confidence I should attach the intention to get some soon, given the snow outside.

The problem of maintaining appropriate confidence—my third question—is contiguous with these problems, but also introduces the question of how end-beliefs relate to means-

beliefs. Suppose that I know how to bring about A by φing, that I prefer that means, and that I intend to bring about A in virtue of having this very belief. To bring about A, I also presumably need a self-referential belief that I will φ, so I form that as well. My confidence that I will bring about A should now depend in part on my confidence that I will φ, which raises the question of how to explain this link.

One potential route to an explanation is by way of the self-reference of the end-intention. One usually arrives at a means-intention in virtue of having an end-intention. The latter's self-reference might therefore add both the arriving at the means, and one's attempts to bring those means about, to the truth conditions of that belief. However, a failure to bring about A by φing does not preclude an attempt to bring it about by φ'ing. When my car gets stranded, I might decide to walk to the store instead. The dual role of credence in a traditional Cognitivist intention makes a muddle of such reconsideration. The theory portrays credence as the source of motivation to reason instrumentally and act, and also (I am presuming) as the basis of technical confidence. When a partial plan fails, one might either try a different way or give up. Trying a different way would seem to be a manifestation of the motivation role, while giving up would likely have more to do with technical confidence in the other options. Given other possible means, the right choice depends in part on the strength of one's motivations, which the credence is supposed to mediate. But the relation of credence to content is fixed at the time the intention is arrived at.

3.2.2 Setiya's Theory

The variant of Cognitivism developed by Kieran Setiya deserves a separate discussion, both because it has distinctive features relevant to technical confidence and because "knowing how" and its relation to intentional action is a focus of his work. According to his theory, an intention is not just a belief but a "desire-like belief"—an attitude combining aspects of desire with aspects of a belief[11]. So rather than just do without a desire component, as in more traditional

11. Setiya is not explicit about just what aspects of a belief this attitude has. His usual characterization is that it is a kind of belief, which he then describe in terms of the conditions under which it can be rationally

Cognitivist theories, Setiya fuses the two into a single attitude.

In making the belief also a pro-attitude, Setiya addresses several of the concerns just discussed. First, the desire for the represented outcome provides a motivation distinct from the epistemic one to have accurate beliefs. Accordingly, that motivation is the ultimate source of the warrant transmitted by the self-referential content of the belief, which his theory retains: you believe that this very belief will bring about the end because you desire that it will. And because the degree of your desire can vary independently of your credence, and ultimately depends on the value you see in pursuing the intention, his theory has a better account of how you can revise an intention. In effect, once your desire falls below a threshold, the attitude would no longer have any force *as* an intention. Even with this change of motivation, however, Setiya also considers planning, including the selection of appropriate means, to be governed by epistemic norms. When you intend that A, you should eventually intend necessary means φ in order to avoid epistemic inconsistency (Setiya 2007a, pp. 666–7).

The change in motivation allows a corresponding change in the requirements for adopting a desire-like belief. The model in Velleman's theory is more epistemic; it emphasizes the overall likelihood of a self-referential belief becoming true once it is adopted. Accordingly, he cites a variety of factors that would weigh for or against an intention being (or becoming) true, including one's motivations and skills, and knowledge of incompatible circumstances. In Setiya's theory, *knowing how* is the main requirement on the belief aspect of the attitude, with beliefs about one's circumstances also potentially factoring in. He says about this requirement (Setiya 2009, p. 135):

> When I am doing φ intentionally as a basic action—not by taking further means—I must know how to φ, and my knowledge consists in the disposition to φ in execution of my intention. (Since dispositions can be masked, this knowledge is consistent with inability, in the simple conditional sense.) When I am doing φ intentionally as a non-basic action, and I know how to φ, my knowledge consists in knowing basic means and knowing how to take those means.

held. However, at least once he says that the attitude "has some features of belief—for instance, it may count as knowledge" (Setiya 2007b, p. 109).

Unfortunately, while these alterations to the Cognitive approach do emphasize technical considerations over others, they do not much improve the answers to the three questions. Separating the motivation from the credence opens up different ways of encoding lower confidence, such as keeping the content of the attitude fixed but lowering the credence in it. However, with respect to appropriate initial confidence, the theory is no more specific about how "knowledge how" and knowledge of the circumstances should be combined into a unified judgment of confidence, a problem that Setiya acknowledges[12]. The previous quote continues:

> There are difficult questions here. How does such local knowledge—knowing how to φ here and now, on a specific occasion—relate to general knowledge how? What is the path from knowing how to bake pies to knowing how to bake this particular pie in this particular oven? Is general know-how just knowledge of generic means, or does it require knowledge how to take those means or a capacity to find and take their instances in contexts that call for it? …
>
> Fortunately, we need not settle these questions in order to address residual puzzlement about practical knowledge, at least to some degree.

Setiya's theory also includes a specific account of the relation of means to end, in terms of beliefs about reasons. He says "To take *p* as one's reason for doing φ is to have the desire-like belief that one is *hereby* doing φ because of the belief that *p*" (Setiya 2007b, p. 46). Suppose that *p* is stopping the main from backing and φ is easing the jib. Then "the belief that *p*" would be the intention to stop the main from backing, which one's intention to ease the jib would refer to as one's motivation. However, I am not sure whether or how *this* relation would support the reassessment of appropriate technical confidence in light of changing circumstances. Setiya's template encodes an instrumental dependence of means on end, whereas the maintenance of ongoing technical confidence requires a dependence of end on means.

12. In most of his papers Setiya also asserts that *knowledge how* is non-propositional. If that claim applied to every skill sometimes described as "knowing how", including knowing how to poison a household, it is not clear how one would *reason* about it in any detail. Absent such reasoning, it is unclear how would one take beliefs about one's circumstances into account. In his most specific paper on the subject, however, I read him as clarifying that only the knowledge-how of basic action is non-propositional in a strong sense (Setiya 2012, p. 296).

3.3 Inferentialism

I have so far discussed three classes of theory. According to the first, an intention is a paired belief and desire, in the second it is a belief only, and in the third it is an attitude with attributes of belief and desire fused into a single state. This summary alone suggests a possible fourth class, in which intention is a desire only. Of course, the simplest version of such a theory would face one of the objections to the simplest pair theory: that we desire some outcomes that we do not intend. An "intention-desire" would therefore presumably need some feature, of form or content, to distinguish it from other desires.

Sarah Paul's Inferentialist theory is of this class. According to Paul (2009a, pp. 11–2), an intention is not a desire, strictly speaking, but a pro-attitude towards some goal that is generally the outcome of a decision to pursue it. What are "inferential" are not one's intentions, but any beliefs about the likelihood that one will realize them. Such beliefs are not *part* of an intention, but inferred from introspective knowledge of it, along with beliefs about one's skills and the circumstances likely to obtain when one hopes to act.

Three factors make it a bit harder to answer the three questions in relation to Paul's theory. One is that it is a relatively recent and less debated view. Another is that she proposes it as an approach to technical confidence[13] compatible with multiple theories of intention (p. 11). Inferentialism is therefore one component of a potential theory of intention rather than a complete theory. The third factor is that Paul largely limits her discussion to high-confidence cases (p. 14, fn. 55). My discussion here is therefore more speculative in comparison to that of the other theories.

In an Inferentialist theory, high confidence is not encoded in an intention but in a belief ultimately derived from it, or more specifically in the credence in that belief. Lower confidence could be represented either in the content of a belief ("I will make A more likely") or with a lower-credence epistemic attitude. Would someone necessarily have, or *should* someone have,

13. She does not use that term, of course.

any such a belief about a given intention? Paul argues against any necessary connection. Whether one should have such a belief would depend on whatever norms govern speculation about one's chances of success, which I read her as leaving open. An epistemic closure requirement is one possibility, and a general interest the outcome of one's attempts to act is another.

To help consider what would make initial technical confidence appropriate, we can first limit the answer to the time just after an intention is arrived at. At that point, the account is similar to the one I offered for belief/desire pair theories. Because the confidence, whether high or lower, is encoded in a "normal" belief, it is subject to the standard epistemic norms, including that the belief be justified[14]. Appropriate confidence will therefore depend on the belief having been properly justified in virtue of warranted beliefs about one's skills, the relevant circumstances, and knowledge of the intention itself.

Strictly speaking, such a justification is only possible once the decision has been been made. The outcome of such a decision, however, will usually depend in part on one's confidence in being able to bring about different potential ends. Paul's theory therefore raises the question of how technical confidence can factor into that decision in the first place. Below I raise what I take to be a more general version of this problem, but at the point the decision is made, the gap seems easy enough to cross by way of some hypothetical reasoning: If after deciding you can reason to your actual confidence, before deciding you can likely reason to your potential confidence.

Any account of ongoing, appropriate technical confidence depends in large part on the relation between end-intentions and means-intentions, an unfortunately Paul says little about that relation. Even in the worst case, however, the two might be linked by way of some further psychology. Suppose you decide to bring about A, and then later decide to φ, but there is nothing in the form or content of your intentions that links the two[15]. On this model,

14. It seems safe to assume the belief must be justified, rather than have some other kind of warrant, if it is ultimately inferred in part from a belief about the intention.

15. Epistemic justification being one example of link between attitudes by form rather than content: a justified belief is linked to a justifying belief, but neither *refers* to the other.

they might still be linked by beliefs inferred from a further psychological fact. I presumably understand that I decided to φ in order to bring about *A* (or to make doing so more likely). In that case my subsequent belief that I am less likely to successfully φ can and should bear on my belief about my likelihood of bringing about *A*. If the link between end and means were instead more explicit, this sort of inference would be more direct. However, in either case, the account of ongoing confidence would be an epistemic one, with changes in confidence tracing to changes in the warrant for beliefs about confidence.

What Inferentialism lacks is an account of how intentions are to be revised in light of changing circumstances. If confidence were only relevant at the time of initial decision, some hypothetical reasoning at that time would be sufficient. The primary argument Paul makes against Cognitivism is that there is no necessary connection between intention and belief (Paul 2009a, pp. 7–9). Regardless of whether that is true, there *is* an ongoing normative relationship between technical confidence and intention. Whether I should continue to have a pro-attitude towards bringing about *A* (as opposed to some alternative plan) depends in part on my continued confidence that I can bring *A* about. By excluding the inferred beliefs from the attitude of intention, Paul also excludes the influence of any subsequent reassessment of technical confidence. The addition of norms to establish this link would make a fiction of the separation between the intention and the inferred beliefs[16].

3.4 Metacognitive Approaches

In contrast with the other theories, Inferentialism also adds a psychological element to practical reasoning. In a pair theory, the template for the belief is something like that φing will bring about *A*, or that it will make *A* more likely. Such a belief is not (generally) *about* any psychological state, it is about an end and a means to it. The self-referential belief of a Cognitivist theory may have psychological content, in that it refers to itself and therefore refers to a

16. Paul also considers a modified version of Grice's theory, in which a belief is inferred from the pro-attitude but both are considered part of the intention. With respect to technical confidence, his view is more like a variant of the pair theory, with a different "template" for the belief.

belief, but the impact of that psychology on other aspects of the theory is minimal. The initial inference from an Inferential intention, in contrast, would yield something like the belief "I intend to bring about *A* "—a belief about an intention. This form of initial inference has some potential advantages. As Paul notes, if you are aware that you often forget to do what you intend, you should have lower confidence in your new intentions (Paul 2009a, p. 15). But it also means that any assessment of technical confidence would depend in part on the application of a theory of mind.

Paul's theory is therefore metacognitive, or at least it is arguably so, given that the concept *metacognition* is broad and interpreted differently by different theorists. The descriptive slogan of metacognition is "cognition about cognition" (Fleming, Dolan, and Frith 2012, p. 1280), and Inferentialism qualifies by that standard. However, another criterion that is sometimes offered is a degree of displacement or indirection from cognition at the "object level"—cognition about things, events, and processes in the world (Nelson and Narens 1990, pp. 125–6). Suppose, for example, that you have what you take to be a memory, but you realize that the "remembered" event very likely never happened. You don't remember any events leading up to it, and have other memories inconsistent with it having occurred. Reasoning about a false memory in this way is a metacognitive process, in that you are taking attitudes *towards* your (perhaps only apparent) memories in order to judge their accuracy.

Whether the slogan and criterion pick out the same thought processes in a given cognitive domain depends on whether that domain *has* a relevant "object level"—on whether routine cognition of that kind only concerns things, events, and processes in the world and is generally free of psychology. The displacement criterion suggests that metacognition is exceptional cognition; it is cognition at a higher level *about* cognition at a lower level. A capacity for it allows the bearer to detect and correct problems that arise in relatively routine thought processes. One normally arrives at a belief by warranting it. Because it rained, you decide that the ground is wet. When one is confronted with inconsistent beliefs, however, one must evaluate chains of justification from a critical perspective perspective, considering each element *as* a belief that might be inaccurate. Seeing that the ground is dry, you might question whether

you should believe that it rained. One can even evaluate one's perceptions in this way, at least to a limited extent. By realizing that the content of a visual experience is improbable or impossible, you might conclude that you are hallucinating.

One way of expressing the tension between the "cognition about cognition" slogan and the displacement criterion is by asking "Why wouldn't there be an object level?" According to Inferentialism, every judgment of technical confidence is based in part on a belief about an intention. I have argued that such judgments are routine, and are therefore not metacognitive in the narrower sense. If that is right, the question becomes why it is necessary to dip into psychology to make such judgments, or *whether* it is necessary. Paul is certainly correct that someone who routinely fails to act on his intentions (and who notices) should take those failures into account when assessing her present confidence. But does that mean that all judgments of confidence must therefore be grounded in psychology, or only that *that* kind of judgment needs to be[17]? Do people, including younger children, routinely apply their sense of how reliably they act in arriving at plans? I decide to move the matchbook from one side of the table to the other, and am confident that I can do so. Do I really need to reason *about* my intention, actual or prospective, to arrive at that confidence, and if so, why?

Evolutionary considerations provide one argument for a generally psychology-free "object level" of intentional action. Many animals with relatively simple decision-making capacities are able to pursue goals, and therefore act in their own interests in some sense. The evolution of the human species likely passed through a stage that included a capacity to represent a goal—we might call these "proto-intentions"—but lacked higher-level reasoning or a theory of mind. If the decisions of these distant ancestors were made at an object level, we might, absent further evidence, also expect our most routine goal-directed reasoning to have the same basic pattern. And if for some reason it has changed, we can rightly ask what subsequent evolutionary pressures prompted that change. This kind of argument is far from definitive, of course. Perhaps those ancestors merely had "proto-intentions" which did not include any

17. To be specific, I am suggesting that taking a tendency to "flake" on intentions into account in this way may itself be metacognitive reasoning in the stronger sense. If it is, then the inferences made in that reasoning would be the exception rather than the rule.

representation of the likelihood of success. Or perhaps the "full-blooded" human decision-making that Velleman argues for is already sufficiently distinct to conclude there is such an evolutionary story, even if we do not yet know its details.

Arguments about what is or is not metacognitive raise many difficult questions, but Inferentialism does not appear to be motivated by that sort of concern either way. Instead, it seems that Paul has arrived at a metacognitive theory through a process of elimination: by explicitly eliminating the theory that intentions are beliefs, and implicitly eliminating any other source of technical confidence. Therefore, although she notes some independent benefits of higher-level cognition, it is the apparent lack of a suitable ground level for practical reasoning that leads her to conclude that it is routine.

3.5 Guidance and Control Theories

The last class of theories to consider are not of intention but intentional action, of what distinguishes such actions from other behavior or movement. Frankfurt (1978, p. 158) argues generally against Causalism, and particularly against its motivating premise that no property can differentiate intentional and accidental behavior at the time of acting. He proposes that what separates these is *guidance*—purposive behavior attributable to an agent[18]. Guidance is not a property of movement. Complexity of movement can suggest the influence of guidance, but the pattern alone is not conclusive. Indeed, even a *lack* of behavior can be the product of guidance. Frankfurt says, "A driver whose automobile is coasting downhill in virtue of gravitational forces alone may be entirely satisfied with its speed and direction, and so me may never intervene to adjust its movement in any way. This would not show that the movement of the automobile did not occur under his guidance. What counts is that he was prepared to intervene if necessary, and that he was in a position to do so more or less effectively." (p. 160). Agents have the capacity to guide, and it is in virtue of an exercise of that capacity that movement (or a lack of it) is intentional.

18. An example of purposive behavior *not* attributable to the agent being an iris contracting in bright light.

This conception of *guidance* is similar to the "cybernetic" theory of *goal-directed systems*, which has also been applied to the problem of intentional action by Berent Enç and Frederic Adams, among others. One of the animating questions of cybernetics is what kind of system can be rightly said to have a goal. If some use of representation (or at least information) is necessary, it is not generally viewed as sufficient, as a system that merely maintains an equilibrium may not qualify[19] (Adams 1986). What counts as a goal might accordingly be applied to the question of what counts as guidance, perhaps serving as a first principle of agency, albeit not the only such principle. The theory has also been adapted as an analysis of what counts as an intentional (as opposed to accidental) realization of a goal (Enç 2004).

The focus of these theories on a kind of productive *feedback*—the application of new information to increasing the likelihood of realizing a goal—makes them especially applicable to the third question of technical confidence. One's chance of realizing an end depends on the chances of realizing one's chosen means, and also on the likelihood that the circumstances are such that those means will contribute to the end's coming about. Awareness of a failure to bring about a means, or of a change in circumstances, can call for a change of plan, or of confidence concerning the end, or both. Therefore, whatever the norms of ongoing assessment of technical confidence are, if one were to "observe" the operation of a system instantiating those norms, it would generally exhibit the kind of feedback that could be also called guidance or goal-directedness[20].

The fact that we have the capacity for productive feedback, and make routine use of it, raises an interesting difficulty for any theory of instantaneous confidence, including initial

19. The goal criterion might therefore also differentiate a contracting iris from other intentional behavior, given that the former might be explained in terms of an equilibrium of light-intensity, signaling from the retina, or other factors.

20. I am speaking loosely here, and not claiming a necessary connection. One can imagine a creature, or just a particularly stubborn person, who never changes a plan after deciding on it while continually reassessing its chances of success. Or a creature or person might build many contingencies into a plan and apply those contingencies in light of changing circumstances without ever reassessing their confidence. My point is that both processes involve similar patterns of feedback, and that actual instances of replanning are often both the result of, and depend upon, reassessments of confidence.

confidence[21]. A simple theory of technical confidence might follow this line of reasoning: An agent comes up with a plan to realize some end that includes some contingencies for the alternative future circumstances she can easily foresee. The agent's technical confidence of realizing the end would start with her confidence in the plan, in virtue of her assessment of the relative likelihood of those foreseeable circumstances. She would then reduce that confidence by a factor that corresponds to *unforeseeable* circumstances. The problem with this approach is that it will generally underestimate the appropriate level of confidence: just because the agent does not foresee a difficulty does not mean she will be unable to circumvent it.

To take these thoughts to a extreme conclusion, why should technical confidence about some end *ever* depend on confidence in tentatively chosen means, up to and perhaps even past the point of acting on those means? The prevalence of discussion of "instrumental reasoning" reflects a genuine pattern in practical reasoning, which is that we often decide on an end based on general technical considerations and defer further planning. So why bother addressing contingencies before they even arise? The most basic answer to this question is that recognizing a circumstance *as* a contingency, relative to a plan, depends on some up-front reasoning. In order to realize an end one must eventually take actual and specific means. If one will not recognize the opportunities to take means, one should not have high confidence. Many plans also involve multiple ordered steps, and the chances of eventual success may depend in part on completing a given step while leaving enough time for the rest.

Pursuing a given means can also be an investment of time, energy, or other resources, and minimizing those investments is an important aspect of planning. Indeed, planning itself takes time and cognitive energy. And according to research I discuss in Section 6.6.2, having a specific plan increases one's awareness of opportunities to act on it. These considerations suggest that technical confidence should depend on a mix of general technical knowledge (including "knowledge how") and confidence in chosen means, with the balance shifting somewhat to the latter as the time to act approaches. And one's means can of course have their own further

21. The first question of the representation of technical confidence lies outside the scope of these general theories.

means, subject to the same balance, which should also weigh on confidence in the overall end.

There may be as many sources of general technical knowledge as there are of general knowledge of any kind. Some such knowledge is clearly inductive. I may doubt my ability to drive to a destination in the snow based on past experience, while being confident on dry pavement for similar reasons. Some theorists have also argued that we sometimes have a trustworthy sense of what we will be able to do not based in reason, at least over a given range of circumstances. I support a version of that view in relation to basic action in Section 6.4.3. A general account of technical knowledge is a project in itself, and outside the scope of this thesis.

CHAPTER 4

Two Stakes in the Ground of a Different Approach

A good theory of appropriate technical confidence might still be grounded in some combination of belief and desire (or Inferentialist-style intention). I take a different approach in this thesis, although one that retains the idea that judgments of technical confidence are a product of reasoning. This combination of claims is unusual. To say that one's confidence in the likelihood of bringing about some goal can be appropriate or not is to say that it is subject to a kind of accuracy. And while it is widely accepted that there are representations other than beliefs that are subject to accuracy, including those directly produced by our perceptual capacities, those are still plausibly *epistemic* in some sense. Belief/desire pair and Inferentialist theories make technical confidence a matter of familiar epistemic reasoning, while Cognitivist approaches change the type of reasoning but still ground technical confidence in an intention-belief. The theory of Chapter 7, in contrast, involves a kind of reasoning that directly supports an attitude that is not epistemic but is still governed by norms of accuracy[1].

My inspiration for this approach comes from Anscombe, and particularly her already-discussed example of the shopper and the detective. But the theory is better illustrated with a simplified case in which the knowing and acting are more "pure"—detecting is also kind of action—and high confidence is more clearly appropriate. So "let us consider" two people, Agatha and Obadiah, standing on opposite sides of a table. They have grown up together and, where relevant, have the same warranted background beliefs and have had similar experiences. On the table are two slices of bread, a knife, and a jar of peanut butter. Agatha and Obadiah

1. I have learned that the prospect of such an attitude and associated norms of reasoning strikes some philosophers as not just novel but counter-intuitive. As someone who has a healthy if not absolute trust of intuition, and a greater trust of counter-intuition, I am not happy about this. All I can ask those that share this sense is to suspend judgment until the end of Chapter 7, and then test your intuition against the specific proposal.

are familiar with knives, jars, and slices of bread, and know that this particular jar contains peanut butter. Agatha picks up the knife and dips it into the jar, getting peanut butter on it. Then she uses it to spread that peanut butter on one side of one slice of bread. Lastly, she puts the knife down, picks up the other slice, and puts it on top of the first one. As it happens, none of the peanut butter is then visible between the slices.

This description alone does not establish that Agatha intentionally made the sandwich; she might have been hypnotized beforehand, or her thoughts about her behavior might be incoherent. Similarly, Obadiah may not wind up knowing that a peanut butter sandwich is on the table; he might have suffered from hallucinations, or just not have paid attention. But it would be quite ordinary if Agatha did intentionally make the sandwich and Obadiah knew what it was. So suppose that Obadiah does observe as Agatha's arms and hands made these movements, and ends up at least *believing* that there is a peanut butter sandwich. When asked why he believes, he says:

> Well, the knife dipped into jar, and got peanut butter on it. Then the knife moved to one of the slices of bread, and the peanut butter got spread on it. And then the peanut butter wound up between the two slices when the other one was put on top. And all a peanut butter sandwich is is two slices of bread with peanut butter between them.

Still putting the question of knowledge aside, what Obadiah describes here is a *justification*—part of a warrant for a belief in which its truth or high likelihood is secured in virtue of other warranted beliefs that are related by a pattern of reasoning. I say "part of" because the premises of a justification—in this case the events that Obadiah witnesses—also need warrant. However, on the assumption that those *are* warranted, and if we accept (partly on the basis of common background knowledge) the validity of his reasoning, Obadiah's belief is in a certain sense appropriate.

As described, his belief is also true, but it might not have been. If someone had replaced the peanut butter with wood putty, he would be mistaken about the type of sandwich. And even when the belief is true, it could fail to be something he knows if its warrant is subject

to a Gettier problem[2]. The narrow sense in which the belief is appropriate is that he has met the epistemic norms for warrant that, when followed, make it so that our beliefs are generally true. Whether or not bad luck leads him to false or coincidentally true beliefs, he has done his part.

A complete description of Obadiah's warrant would be a significant undertaking, because it would need to account for each relevant background belief and how it is warranted. Also, some of his reasoning has to do with how physical objects exist and move through space, which may be warranted in virtue of specialized cognitive capacities. But his case is fairly typical from an epistemic perspective. Different epistemologists would describe aspects of Obadiah's warrant in different ways, but few would be flummoxed.

What makes Obadiah's explanation relevant to the problem of action is its similarity to one that Agatha can offer *before* she acts. Suppose that, before she picks up the knife, Agatha intends with high confidence to make a peanut butter sandwich. Asked then "How will you make it?", she says:

> I will dip the knife into the peanut butter jar, so it will have peanut butter on it. Then I will move the knife to one of the slices of bread, and spread the peanut butter on it. Then I will pick up the other slice and put it on top of that one, so that the peanut butter is between them. And because all a peanut butter sandwich is is two slices of bread with peanut butter between them, I will have made one.

One way to characterize this explanation is as a description of means—of other things that Agatha plans to do in order to further her sandwich-making end. But it is not only that. The last "step" is, after all, purely conceptual. According to some theories of action identity, her placing peanut butter between two slices of bread and her making a peanut butter sandwich would be the same act described in two ways. Nevertheless, if Agatha behaved in exactly the same way, but had no concept of *sandwich*, we might be reluctant to say that she *intentionally made a sandwich*. How she conceives of her behavior seems relevant to the question of what she does intentionally.

---

2. See Chapter 10.

Her statement is also more than just a description of means because of how its parts relate to one another. Those relations may be more familiar when expressed in the past tense, as in Obadiah's answer, but her language is similar enough to suggest that it is based on similar reasoning. If we wished, we could re-write both statements to refer to specific times, with the caveat that hers would not be as exact as his[3]. And although she attributes the events to herself throughout, she could instead start with "I will make it the case that ..." and use language even more like Obadiah's. Whatever the phrasing, both statements assert that certain events, given the broader situation, add up to the making of a peanut butter sandwich.

4.1 The First Stake: Shared Patterns of Reasoning

On the assumption that Agatha has appropriate confidence that she can bring about each of her means, why *shouldn't* her reasoning be the same as Obadiah's? After all, one could describe her as "justified" in thinking that she will make a sandwich given her explanation, and assuming that she does not change her mind. Perhaps Agatha's and Obadiah's descriptions are so similar because the mental states and relations they correspond to are similar. That is, whatever chains of reasoning that stretch from justifying to justified beliefs in Obadiah's thinking could be mirrored in those that stretch from means to ends in Agatha's, justifying her high confidence in the latter.

This is a stronger and more specific claim than that they have closely analogous *reasons*. In terms of the simile of a formal proof, their having analogous reasons for A would be like each having proved A from the same premises. In that case, the individual lines of Agatha's proof might differ from those of Obadiah's, reflecting differences in argumentative "strategy". I am instead claiming that their justifications could be as if each line of their respective proofs, and the rules that relate those lines, are the same. The differences would therefore be not of "syntax" but "semantics"—in the attitudes that Agatha and Obadiah take towards their "formulas". Not

3. Davidson (1963, p. 6) argues that an agent's intention cannot pick out particular future events the way that an observer's perception can refer to past events. He might therefore object to this aspect of the comparison. I argue against this view in Section 11.2.2.

every attitude would differ, given that both justifications clearly depend in part on a number of background beliefs. A "technical justification", whatever that winds up being, would therefore need to incorporate both means-intentions and beliefs into the justification of an end-intention, while preserving the *patterns* of reasoning that could also justify a belief.

That Agatha and Obadiah can use the *same* pattern is of course an artifact of a carefully chosen case. The shopper's justification might overlap with that of the detective, but they would probably not overlap completely. It is doubtful each would mention all of the same events, for example. The shopper might pay the most attention as he puts each item into the cart, while the detective records each as it is rung up. The value of a shared-pattern case comes from contemplating a theory of technical confidence that *can* use patterns of reasoning normally associated with epistemology. If the same pattern applies in such cases, the same basic strategy might work more generally, albeit with different patterns appropriate to the situation.

The principle that ends and means are related by the same patterns of reasoning that relate justified and justifying beliefs can therefore be the first stake in the ground of a theory of technical confidence.

4.2 The Second Stake: Technical Confidence is the Analogue of Credence

A second principle follows quickly from the first. The "conclusion" of Obadiah's reasoning is (his belief) that there is a peanut butter sandwich. His justification is what makes that belief, and more narrowly its high credence, appropriate. Credence is, in effect, *epistemic confidence*. The corresponding attitude in Agatha's reasoning is her intention to make a sandwich or, removing agent-attribution from the content, her intention that there be a peanut butter sandwich. Given that she is confident that there can be one, how might her confidence be encoded?

All of the theories of last chapter are, it appears, already ruled out by the first stake. Cognitivism does not apply because Agatha's intention is just that there be a sandwich. Its content

has nothing to differentiate it from a typical belief, and therefore her simply *believing* that there will be one would negate her own agency over the matter. If we try to address those problems with some self-reference, her reasoning would be left incomplete at best. To *justify* additional self-referential and causal conditions would require adding corresponding content to the means-intentions to encode what caused them, and then some additional reasoning would be needed to capture the relation[4].

Inferentialism would only apply to Agatha's reasoning if it were to conclude in a belief *corresponding to* her intention, rather than the intention itself. But then her reasoning is of the wrong kind, because it does not take a belief *about* that end-intention as a premise. Obadiah's explanation does not refer to his own cognition, although he could have made a different, more psychological case. Instead of just discussing the facts "on the table", he might have explained in terms of what he "saw". Given that the question he answers is about why he *believes*, this would be a natural choice. Even then, however, his connections of reasoning would still not be psychological. "Seeing" is factive, and his "move" from seeing the slice placed to the existence of a sandwich makes at most a trivial reference to his visual capacities.

Finally, the belief/desire pair theory also does not apply because the belief in its template is general, while Obadiah's reasoning concerns particular events.

On the assumption that Agatha uses Obadiah's pattern of reasoning, it appears that her confidence must be in her intention to make a peanut butter sandwich, which is the attitude that corresponds to Obadiah's belief that there is one. That attitude is not a belief, and cannot have different *content* without a corresponding change of reasoning. The remaining option is to build technical confidence into the attitude of intention itself, as credence is built into belief. As with lower credence, lower technical confidence might be encoded either in the content of a formally high-confidence attitude ("I intend to make it more likely that …") or in virtue of a distinct attitude that associates a lower degree of confidence with an unaltered content.

4. The result might be something like "And because all a peanut butter sandwich is is two slices of bread with peanut butter between them, and my intention to pick up the other slice and put it on top of this one was caused by (or in virtue of) this very intention, I will have made one."

I have already argued that technical confidence is central to practical reasoning. The proposal that it integral to the attitude of intention implies the stronger claim that one cannot intend without some sense of one's chance of success. That sense could be just as indistinct as the vaguest epistemic "hunch", for which one can offer no particular odds. But it would be a necessary aspect of intention, in contrast with, for example, desire. This claim is different from the Cognitivist premise that intention is necessarily associated with belief. Technical confidence would not be in virtue of belief, but of a different attitude that "measures" the chance of a voluntary outcome: *Given* one's commitment to the intention—that is, contingent on one's continuing to intend—one is confident (to some degree) that what is intended will come about. Setting aside this particular theory of how technical confidence is encoded, I take the weaker claim that we have some sense of confidence in *every* intention to be plausible[5].

The principle that technical confidence is to the attitude of intention as credence is to belief will therefore serve as the second stake in the ground of a theory of technical confidence.

4.3 A Sketch

The two principles together suggest a generalization and an approach. The appropriateness of Obadiah's credence is in virtue of his epistemic justification, and by stipulation the appropriateness of Agatha's technical confidence is in virtue of what I have called her technical justification. The general norm that Obadiah meets is that his beliefs be epistemically warranted. By analogy, we might seek a theory of "technical warrant" modeled on epistemology, but adapted to account for appropriate technical confidence. Given the (rough) relation of Obadiah's perceptual beliefs to Agatha's intentions to perform basic actions, that theory might include a kind of *technical entitlement* to intentions for such actions. In short, rather than "borrow" the attitude of belief from epistemology to explain intention, we might instead adapt the theory of epistemic warrant to account for confidence in voluntary commitments, resulting in a theory of warrant that serves the opposite direction of fit.

5. I briefly return to this subject in Section 7.6.

As I argue in Section 3.5, technical confidence in a given intention should depend in part on one's chosen means (if any) and also on general knowledge (which might include "knowledge how"). A complete theory of confidence must therefore account for both sources and the balance between them. However, as I note in Section 1.3, I am deferring the problem of partial plans and replanning as future work. In this thesis I therefore discuss technical confidence almost entirely in terms of "complete" plans.

CHAPTER 5

An Epistemic Model

In the last chapter I proposed to account for appropriate technical confidence on the model of epistemic warrant. Before I can take this approach, however, I must overcome two challenges. The first is that there are a vast number of topics that are at least partly related to the subject of epistemic warrant, and discussing even a large fraction of them would overwhelm any discussion of intention. The second is that the subject encompasses many incompatible theories and many ongoing debates, which means I must either present the various sides of those debates or choose some sides over others.

The first challenge has an easy solution: I can discuss only those issues that I make use of in the technical theory[1]. (This is not to say that other issues have no technical counterpart—I have also had to pick what topics to discuss in the rest of the thesis.) The content of later chapters also determines the *depth* of the discussion in this one. For example, because what I call "patterns" of justification are common to both realms, I include only a high-level characterization of the form of epistemic justification and a very brief survey of the most common types.

With respect to the second challenge, the theories I present are largely those I personally, although sometimes tentatively, hold. When I have no strong opinion, I present the view I take to be most popular. This approach allows me to address technical questions in relation to my personal understanding of the analogous epistemic questions, rather than trying to do so by

1. Almost all of my discussion of belief and epistemic warrant, short of knowledge itself, is in this chapter. The exceptions are the discussions of the "behavior" of belief in Section 6.6.2 and probabilistic consistency in Section 8.3.2. I address the larger subject of (reflective) knowledge in Chapter 10. Because this chapter primarily serves as a reference, it is divided into more sections and has less transitional material.

way of views that I do not entirely understand or that I mistrust. However, it may also raise the concern that I have cherry-picked the epistemic positions to make the technical theory "work out". My work in action theory has probably changed some of my views about epistemology, but none of these changes were drastic, and in every case I think I can defend the result from epistemic considerations alone. And as I note in Section 1.4, many debates on the epistemic side can be "translated" into analogous technical debates. The goal of this thesis is not to settle all debates, but to provide enough context for their translation.

5.1 Representation and Accuracy

The primary subject of this chapter is the attitude of belief and the norms that govern that attitude. Belief itself can be difficult to describe because it is such a basic element of human cognition. To make that difficulty explicit, I will begin by attempting to describe belief in terms of even more basic concepts.

A belief is a type of *representation*. A standard gloss on representation is that it is the class of things that are *about* things. A representation is said to have "intentionality", and each likely either consists of or is closely associated with some *meaning* or *content*. However, I doubt that these explanations are as helpful as considering the conceptual connection between *representation* and *accuracy*. A representation can be thought of as something to which the concept of accuracy can apply. This is not to say that any representation *can be* accurate. Some, such as contradictions, are inherently inaccurate. Still, to even ask the question "Is this accurate?" of something, one needs a sense of the factors could make it accurate, or that rule out it's being accurate. Those factors are, speaking very roughly, what a representation is "about". Different systems of representation can represent different factors, but the overall domain of representation includes everything that can be measured, referred to, or speculated about. Whatever lies beyond what can be represented lies beyond understanding, because understanding is in virtue of representation.

5.1.1 Mental Representation

Some representations are *mental*. The line around what is "mental" is increasingly controversial, so here I will just say that mental representations are those that are directly participate in the operation of the mind, which is in turn intimately related to the operation of the brain. A pie chart of a university budget printed on paper is a representation, but not a mental one. Someone looking at the pie chart might, in virtue of seeing it, form corresponding mental representations of various types. One type is *perceptual*; she *sees* how it *looks*. Perceptual representation is a difficult and only partly understood subject, but in general, the domain of what a given perceptual capacity can represent is closely related to its associated sensory system. Visual experience, for example, consists of representations suited to what information can be derived from stimulation of the retina, given the shape of the eye and the operation of iris and lens.

Other representations, including beliefs, are *conceptual*. Looking at the pie chart, someone has a rough idea of what portion of the university budget is spent on building construction and maintenance in virtue of a conceptual representation.

5.1.2 Propositions and Our "Attitudes" Towards Them

Humans may form several kinds of conceptual mental representation, but the most widely discussed are *propositional attitudes*. Much of the theory of these "attitudes" comes from philosophy of language, and in particular what are called "propositional attitude reports". Suppose I decide that Obadiah believes that there is a peanut butter sandwich, or that Agatha desires that there be a peanut butter sandwich. In either case (and setting aside some grammatical details), I attribute a mental representation to a person. The content of the attributed representation is whatever follows "that", which can, from a grammatical perspective, be any declarative sentence. The way that we speak about these representations does not seem to depend on the particular language they are expressed in. To translate a propositional attitude report from English to French does not imply that its subject is thinking in French. That suggests a certain

generality of meaning, in that what can be stated in a language can also be attributed as the content of someone else's representation. The domain of what can be represented propositionally closely corresponds to the domain of what can be stated in language.

The accuracy of a proposition has a distinctive all-or-nothing character; each is either true (accurate) or false (not). There are exceptions (or perhaps *apparent* exceptions) to this rule, which arise from difficulties such as *vagueness*. In general, however, the various uses of propositional representations depend on this sharp line of accuracy. There may be no bright line of moisture between *damp* and *not-damp*, and therefore no precise answer to my question "Is that towel damp?" However, when "that towel is damp" is not clearly true or false, the situation does not call for a third degree of accuracy, but a different, more specific question[2].

Propositional attitude reports are so-called because they specify not just the propositional content that is attributed but also some attitude towards it: *belief*, *desire*, *intention* (perhaps), and so on. Each of these attitudes corresponds to certain norms the propositional content should conform to, and roles it might play in further cognition.

5.1.3 Accuracy-dependent Cognition

I have said that all mental representations are the sort of thing of which one can ask "Is this accurate?", but the relevance of this question differs for different representations, including different propositional attitudes. When we imagine or speculate, we do so *about* something, and therefore in virtue of representing that thing. But the accuracy of such representations is largely irrelevant to the process of imagining or speculating. Similarly, when you desire something you will—you expect, or hope—be pleased if the desire winds up being accurate, but the role of desire has no direct relation to accuracy[3].

2. The shift from one question to another can be so quick and natural as to almost disappear. Suppose that I ask if the towel is damp, and on seeing you hesitate I grab it and use it. It might seem that I acted in virtue of some third degree of accuracy. But this is doubtful. It is more likely I acted on a slightly different standard than I asked about, such as "Is the towel *clearly* damp?"

3. I would say the role of a desire has an indirect relation to accuracy, in virtue of it being a candidate for intention.

However, some types of representation carry an implication of accuracy, and the roles that representations of those types play in cognition are only appropriate (in some sense) given their accuracy. Suppose, for example, that you have a visual experience of a glass of iced tea on the table in front of you. In virtue of cognition "downstream" from that experience, you might try to pick up the glass and drink from it. That attempt is only in your interests when the experience is accurate—when it represents an actual glass on the table. The cognition associated with accuracy-dependent representations conforms to this pattern: Processes (which can include conscious thought) that make use of such representations presume their accuracy (to some degree). Processes that produce such representations are such that what they produce are generally accurate under typical conditions, where what is "typical" might be described in terms of the environment through the evolutionary history of the processes.

The most widely discussed accuracy-dependent mental representations are perceptual representations and beliefs. I will refer to members of the class, the processes that produce them, and the norms they are subject to, as "broadly epistemic".

5.1.4 Belief and (High) Credence

In light of these distinctions of representation, a belief can now be characterized as:

1. A mental representation
2. in the form of a propositional attitude
3. that carries an implication of truth.

That is, belief is at the intersection of propositional representation and accuracy-dependent cognition. There may be other members of this class, such as *denial*—not in the sense of a type of utterance, but as the attitude towards a proposition that it is false. However, the study of formal proof systems shows how reasoning that makes use of different accuracy implications can be "translated" into belief-based reasoning. Given the long tradition of talking of all such reasoning in terms of beliefs, I continue it here for the sake of simplicity[4] and will refer to

4. What other overall conditions of propositional accuracy might there be? Some lines of a natural logic sub-

beliefs and their associated norms and processes as "narrowly epistemic" or just "epistemic".

According to the general pattern of accuracy-dependent cognition, broadly epistemic representations should generally be accurate in typical conditions. The narrowly epistemic version of this claim is that beliefs should generally be true in typical conditions. A realistic version of this claim would need some caveats that lie well beyond the scope of this thesis. I will assume that to the extent humans typically form some types of false belief, those types can somehow be distinguished from other types, generally speaking.

5.2 Justification

Given the generality of potential propositional representation and the diversity of actual beliefs, it is remarkable that beliefs can typically be true. Much of that diversity corresponds to our capacity for epistemic reasoning, which "produces" not only new *beliefs*, but *justifications* for those beliefs in virtue of other beliefs. Epistemic reasoning that produces a justification—a process also called "justification"—makes use of existing beliefs to arrive at new ones. The final step in Obadiah's reasoning, for example, relates a belief about what constitutes a peanut butter sandwich—two slices of bread with peanut butter between them—to a belief that something of that constitution is on the table. In virtue of those beliefs, he concludes that there is a peanut butter sandwich on the table. As with most cognitive processes that make use of beliefs, justification relies on their being true. If Obadiah had been mistaken about what was between the slices, he would be mistaken about there being a peanut butter sandwich. Justifications are therefore sometimes said to "preserve truth": On the assumption that the justifying beliefs are true, it is necessary or highly likely that the justified belief is true.

The de facto exemplar of justification is a logical deduction, as when someone reasons to a belief of the form Q (that the ground is wet) on the basis of having warranted beliefs of the forms $P \rightarrow Q$ (that if it is raining the ground is wet) and P (that it is raining). The content of

proof (which typically correspond to hypothetical reasoning) do not have a definite truth value within the proof but are still subject to certain accuracy conditions relative to other lines. The general success of natural deduction systems over axiomatic systems suggests a similar arrangement of human hypothetical reasoning.

these two premises are related such that if they are both true, this conclusion must be true. However, this necessary relation of truth is not common to all forms of justification. Suppose I believe that my car is parked in front of my building because it was there ten minutes ago and I have not moved it since. The truth of my car location belief is not guaranteed by any combination of other beliefs that I have, and I should not be certain that it is correct. Even so, it is justified. A justification only needs to make the truth of the belief highly probable. Just how probable, and whether that probability depends on the content of the belief or what further use is made of it, are matters of dispute.

Some of my reasoning about my car is *inductive*—based on a regularity of past events, I reason that the regularity will continue in the future. I may also have thought explicitly in terms of the chances of the truth of my conclusion, an example of *probabilistic* reasoning. Beyond these basic features, there are finer categories of justification, and various proposals for the normative requirements that govern them. However, as I note at the start of the chapter, those details are unchanged in the technical theory, so I need not discuss them in detail.

My claim that epistemic reasoning *produces* a justification, in contrast, is relevant and needs further explanation. One can imagine a creature with some of our reasoning capacities but that arrives at new beliefs (or belief-like attitudes) without forming a justification. Once it adopts a belief, it would retain no "record" of its production, or reasons for believing it. If you asked the creature why it believed there is a peanut butter sandwich, or its car is parked out front, the best it could do is try to construct a new chain of reasoning for it. Humans are not like this; we can usually provide reasons for what we believe.

There is controversy over the extent to which reasons offered in explanation of a belief correspond to the reasons used to arrive at the belief. Not all reasoning is conscious—we have too many beliefs to have arrived at all of them consciously. And sometimes the reasons we offer for a belief are implausible, and they can be better explained in terms of motivations that we only partially recognize. Even in such cases, however, there is a normative expectation that a belief has *some* justification. You can ask of a given justified belief, "Why do I believe this?" If you cannot answer, you should drop the belief. If you can, the answer takes the form of

reasons, which correspond to beliefs that could be evaluated in a similar way[5]. Justification is the basis for a range of important metacognitive capacities, including the capacity to resolve inconsistencies[6]. It is also what allows us to communicate our reasoning, so that I might believe what you believe for the same reasons that you believe it.

What, then, is the *form* of justification? How is it "encoded"? A step of reasoning can certainly be represented as a further belief, given how we can talk about our reasoning ("I believe that P because I believe that Q and that R."). But any proposal that such steps are generally encoded as further beliefs is dubious at best. Epistemic reasoning is relational, it accounts for beliefs in terms of other beliefs. If every such step were a further belief, the relation between P and "I believe that P because … " would still need some further explanation. My own preference is to model the encoding of justification as explicitly and minimally relational. It consists not of additional *content* so much as a kind of inter-propositional "glue", associating one belief with others.

5.2.1 Physical Reasoning and Sequence

An important subset of epistemic reasoning relates to issues that are broadly "physical": time, space, the behavior of objects, and so forth. Just how or even *when* we reason about our physical environment is a complex problem. I do not need to provide a detailed model of reasoning about space and objects here, but I do need to say a little about time, and in particular how we reason about *sequence*.

Consider Obadiah's belief that peanut butter was spread on the slice of bread. To function as part of his justification, it cannot be an isolated belief about a particular spreading. Obadiah must also be aware of the relation in time or sequence between the spreading and the other events of sandwich-making. If the second slice were placed on the first *before* the spreading, his reasoning would not be valid. Just how these relations of time or sequence are encoded in

5. I say "correspond" in the hope of side-stepping the debate over objective and subjective reasons.

6. See Section 5.6.

mental states is a murky matter that I do not want to go into. What is important is that they are encoded somehow, and many kinds of reasoning depend on the relative timing of different events.

5.3 Entitlement

Not all of our beliefs are justified in the sense just discussed, even though they are of a sort that are typically true and can count as knowledge. There is an obvious formal reason why we must have unjustified beliefs. If every belief were justified in virtue of other beliefs, then our chains of justification would either be infinitely long or circular. Infinite chains cannot be held in finite minds, and circular reasoning does not typically produce true beliefs[7]. However, the formal problem provides no insight into what types of unjustified belief are typically true, and why. All it tells us is that there must be such beliefs.

It seems safe to say that the *general* problem of unjustified but appropriate beliefs can have no philosophical solution. To ask why it's is appropriate for Obadiah to have the belief that P is to ask for an explanation, and therefore for reasons that also need explanation. Not everything can be explained, if for no other reason than the formal one. The analysis of unjustified beliefs is therefore on the safest ground for those that can be explained, and perhaps even justified in a sense. The idea is: Take beliefs we have good reason to think have been typical among humans, and that non-specialists (including children) still typically have, and explain in scientific or analytical terms why they are typically true. Such explanations will still rest in part on further unexplained beliefs, but are no worse off in that respect than any other.

This approach is now common in epistemology, and one of its pioneers is Tyler Burge, who prefers the terminology I have already used in earlier chapters: the normative requirement on beliefs is that they be *warranted*. Warrant divides into *justification* and *entitlement* (Burge 1993, p. 458). The category of entitlement is a bit of a catch-all. We are entitled to certain beliefs in

7. Anyway, a closed circle of reasons could only be arrived at all at once, and therefore not in virtue of a "truth-preserving" process.

virtue of the reliability of the processes that produce them, even though we may not be able to provide reasons for them, or the reasons we can give do not justify the specific contents of those beliefs. An explanation of entitlement must therefore be in terms of the relevant cognitive processes. The two discussions of this section, concerning entitlement to perceptual beliefs and to beliefs derived from testimony, are simplified versions of Burge's views.

5.3.1 Perceptual Entitlement

Part of Obadiah's explanation of why he believes that there is a sandwich is in terms of events that he saw. Humans have perceptual capacities that produce generally accurate representations of objects and events around us. The visual system, for example, produces accurate representations of (in a narrow sense) what the eyes are pointed at at the moment, and (in a broader sense) of the environment in front of a person's head, in virtue of eye movement and visual memory. With no obvious effort, humans can also derive accurate *beliefs* about their environment from these perceptual representations. Obadiah's seeing the knife approaching the jar[8] is in virtue of his visual system producing a representation of the knife in movement, and his belief that the knife moved to the jar is derived from that representation. The capacity to form perceptual beliefs can be seen as an informational bridge from accuracy-dependent representations in a sense-specific "format" to representations in a propositional format. If there were no such bridge, we could not form further beliefs about what we perceive.

However, neither the accuracy of our perceptions and perceptual beliefs, nor our "sense" of that accuracy, depend on our understanding the capacities that produce them. Our distant ancestors formed beliefs about what they saw without being able to justify them beyond saying "I saw it", and young children do the same. One way of explaining this cognitive "arrangement" is in terms of natural selection. In the past, random mutations to the genetic "blueprint" of a sensory system occasionally improved the accuracy of its representations. That improvement then contributed to the competitive success of the organism, and that of the descendants who

8. Here I set aside some complications of object identification from past experience.

inherited the genes. Given that people undoubtedly do form beliefs based on what they see and hear, and those beliefs are generally considered to constitute a form of knowledge, it is now a common epistemic theory that we are *entitled* to them.

Just what features of cognition are required to explain epistemic entitlement are a matter of dispute. At the simplest end of the spectrum of explanation are forms of Reliabilism, according to which the general reliability of a perceptual system accounts for the entitlement. Burge argues instead that although entitlement implies reliability, more of the evolutionary history of a cognitive system, and its history of use in the life of an individual, is needed to determine both the meaning of our beliefs and whether the system functions well or poorly in a given instance (Burge 2003, pp. 507–14). Entitlement to a perceptual belief therefore depends, in his view, on the functional norms that govern the relevant perceptual system (p. 530). The theory is compatible with a capacity producing consistently inaccurate representations under specific circumstances, as with visual illusions. Absent contradictory evidence, one is entitled to one's perceptual beliefs even when they are not true.

5.3.2 Entitlement to Testimony

Burge also argues that we have a "default" entitlement to adopt beliefs that are expressed by other people. This entitlement is not absolute; it does not sanction trust in everything we hear. When you have evidence that the speaker has lied under relevantly similar circumstances, or that she is prone to forming inaccurate beliefs, that evidence can and should count against trusting what she says. The entitlement is *prima facie*. It is the basis of credulity in children, and is less often the deciding factor in adults only because of the extensive knowledge that adults typically have about social practices and the interests of individuals (Burge 1993, pp. 467–8).

Although one can often offer reasons for what one believes on the basis of testimony, those reasons are not like those of justification described in the last section. When deductive or inductive reasoning justifies a belief, it supports the *particular* content of that belief. A justification of this kind for the belief that P ("it is raining") cannot also properly justify the

belief that not-*P* ("it is not raining"). The reasons one might offer for trusting a person's testimony are not like this. If you have evidence that I am an expert on snails, what I say could support your believing that snails are omnivorous or that they are not omnivorous. If, on the other hand, I explain my reasons for believing that snails are not omnivorous, although you might have to trust some of my premises, you would have your own justification for the conclusion. Belief on the basis of testimony therefore marks a kind of division of epistemic labor.

An explanation of entitlement to testimony must address two concerns. The first is the question of why expressed beliefs will generally be true given an assumption that people generally make statements in good faith—that is, assuming that people generally say what they believe. This amounts to the question of why a person's beliefs will typically be true, which the requirement for epistemic warrant is supposed to answer. When the speaker is entitled to what she believes by the earlier testimony of another speaker, the chain of trust can still trace back to someone with epistemic warrant for the belief that is specific to its content.

The second concern has to do with the circumstances under which we can and cannot trust people. This sort of trust can also be explained in terms of evolutionary fitness. On the assumption that representing the world accurately contributes to fitness, an epistemic division of labor has obvious benefits to group members by increasing the extent of what they represent. It also allows for "epistemic specialization", where one group member might understand a subject in a way that others do not, and in some cases could not. Those representations will tend be accurate as long as the group tends to express accurate beliefs to children in the group, and educate children about potentially overriding individual interests and social norms. This core of trustworthy testimony, including advice on when to trust, can be the basis of more sophisticated judgments of potential deception[9].

9. Burge himself argues for a general, if *prima facie*, entitlement to testimony, while noting that "this issue is more complex than [he] can see through now" (Burge 1993, p. 475). He grounds his view not in an evolutionary argument but in the teleology of reason, a position he maintains in his more recent Postscript (Burge 2013, ch. 11).

5.4 Lower-credence Epistemic Attitudes

Sometimes we are not confident that a proposition is true but still have some evidence for it. Obadiah might suspect that the substance on the knife is peanut butter based on how it looks and what Agatha is doing with it, but also worry it could be something else, such as almond butter. Even when one has less than high confidence about the accuracy of some representation, the chance that it could be accurate, when supported by evidence, can still play a useful role in further cognition. Obadiah can reason from the chance that peanut butter is on the knife to the (roughly equal) chance that there is a peanut butter sandwich on the table, and then to the chance that his daughter, who took a bite of the sandwich, will suffer an allergic reaction.

That we reason this way raises the question of how we represent lower credence propositional content. This question already arose in Chapter 3 in the context of considering how technical confidence would be represented in various theories. There I left the answer open, but here, and in the remaining chapters, I treat such content as being represented in virtue of an attitude distinct from belief that I name (for lack of a better term) "suspicion". As belief associates a high degree of credence with some proposition, suspicion associates a lower degree with one[10]. The degree might be specific, such as attributing a thirty percent probability, or vague, as when you have some evidence for a proposition and some against, with no specific idea of how to synthesize these into even an approximate number[11]. Suspicion could overlap with belief in the range of high credence. As I am using the terms, the attitude of 99:1 odds that

10. There are two straightforward alternatives to this approach. One is to make belief the only broadly epistemic propositional attitude and make lower confidence a matter of content. For example, Obadiah's lower confidence could be a belief that "There *might be* a peanut butter sandwich on the table." or "There *is a forty percent chance* that there is a peanut butter sandwich on the table." The other alternative is to make "suspicion" the only narrowly epistemic attitude, and model what we call a "belief" as a suspicion that happens to associate sufficiently high confidence with the proposition in question.

11. In some cases, as with coin flips, you have the information needed to calculate a specific probability. In others you may just be ignorant of the chances. But a lack of information can conversely be modeled in terms of probability, as when you decide to treat an event of unknown likelihood as having even chances. So while a sense of credence and a sense of probability are not the same thing, they can to some extent be exchanged with one another when reasoning.

there is a peanut butter sandwich is a suspicion, whereas the attitude that there is a peanut butter sandwich "but I could be mistaken" is a belief.

Just as a belief should have warrant that supports a high degree of credence, a suspicion should have evidential support for its associated degree of credence, however vague or specific. This can also be expressed as a requirement for epistemic warrant, with the understanding that different attitudes call for different forms of warrant. In the examples just discussed, the suspicions are supported by reasoning, and we might therefore say they are justified. There are likely as many forms of lower-credence justification as there are high-credence forms, with a great deal of overlap. Seeing the same person walk by your door every Monday morning for four weeks may not give you enough inductive evidence to believe she will walk by it next Monday morning, but absent other information, you can rightly suspect that she will. Probabilistic reasoning has also been given a number of formalizations, including the Bayesian model, which provide a general approach to updating credence in light of new evidence.

I will not delve more deeply into the subject of lower-credence justification than I have into its higher-credence form, for the same reasons. I do need to clarify, however, that the distinction is really one of end result rather than of "complete" justification. Obadiah's suspicion that there is a peanut butter sandwich (relative to his almond butter doubts) is justified in part by bread-slice *beliefs*. A suspicion can also contribute to the justification of a belief. Suppose I rightly suspect that there is a three in five chance of my motion-triggered camera photographing a raccoon in a given week. I might then justify a belief that it will have photographed a raccoon after twenty weeks, based on my suspicions about each week and a probability calculation.

5.4.1 Lower-credence Entitlement

Must a lower-credence attitude be justified to be appropriate? Morrison (2016) argues that "our perceptual experiences assign degrees of confidence." Even on the assumption that visual representation is limited to depth and color, he notes, we are not always certain about those aspects of experience—it is possible to be uncertain about the color of a tablecloth or the

distance of a ball (Morrison 2016, p. 24). However, Morrison argues that higher-level aspects of visual experience provide even clearer examples of degrees of perceptual confidence. These aspects include shape (pp. 32–3), category (such as alphabetic letter) (p. 17), and recognizing a particular object (pp. 15–6). Each of these aspects of perception include the possibility of uncertainty. You can see a person and have no idea who they are, or you can see a person as your friend Isaac, but you can also see a person as *possibly* your friend Isaac. Or, in a common scenario (p. 17):

> While you're at an optometrist's office she asks you to identify letters on an eye chart with the help of a series of lenses. At first, your experience will be too blurry to give you much confidence that a particular letter is an **E** rather than an **F**, **B**, or **G**. But as she improves your visual acuity you'll report increasing confidence.

If Morrison is right that perception admits of degrees of confidence, then it would be very strange if perceptual beliefs that directly correspond to less confident visual experience were based on independent reasoning. If you can *experience* "maybe an **E**", then it would be most natural for your propositional suspicion "that is an **E**" to be derived from that experience alone. Such a suspicion would be no more or less justified than the perceptual belief "that is an **E**" is derived from a high-confidence perceptual representation. Following the analogy with high-credence warrant, the suspicion would be in virtue of a lower-credence entitlement.

I largely agree with these arguments, while recognizing how they could be controversial. However, the claim that there are lower-credence entitlements is weaker, and should be less controversial. According to the alternative view of "post-perceptual confidence", in which degrees of confidence are limited to belief, the idea that "post-perceptual suspicions" would be epistemically justified in the relevant sense seems difficult to support. If the relevant standard of justification is the availability of *reasons* to consciousness, Morrison's examples fail to qualify on either view. One can certainly try to consciously reason to the same suspicion, just as one can explain in terms of definitions why what one perceives has the shape of an **E**. However, the same considerations that support the division between justification and entitlement in high-credence cases apply just as well across the spectrum of credence.

The case for lower-credence entitlement is easiest to make in the realm of perceptual belief,

but the idea can be applied more generally. Take, for example, the argument that a "default" entitlement to testimony contributes to group fitness in virtue of a division of epistemic labor, including cognitive specialization. In Burge's view, that entitlement is limited by counter-reasoning alone, which would leave some group members (such as children) particularly vulnerable to being misled—a concern that is admittedly partly addressed if children are generally protected from out-group influence. However, it could be further accounted for in terms of a cognitive bias in children against who they take to be outsiders. Such a bias would apply on the border between understanding what someone has said, and adopting it as one's belief, and would take the form of a variable credence entitlement. A child (or adult) might then simply distrust who they take to be an outsider, or at least apply some doubt to what outsiders say, even though she could not explain why she does so. In this way, transfer of testimony would generally track group membership until a person has learned when to trust "outsiders", and not to trust "insiders".

5.5 Warrant for Future States and Events

Most of the examples of warranted belief in this chapter represent past or ongoing states and events. That focus reflects a real asymmetry between the number of high-credence attitudes people generally have about the past and present versus those they have about the future. Given that (proper) intentions are necessarily directed at present or future states or events, I will briefly consider the source and significance of this asymmetry.

It is sometimes said that we know more about the past than the future because the past is "settled" and the future is not. But physical theory, at least at the macro level, provides only limited support for that explanation. There are some aspects of physics that are asymmetric in time, such as the thermodynamic law that the total entropy of a system never decreases with time. But many physical laws are time-neutral, and would therefore allow similar sorts of knowledge about past and future. In virtue of astronomical models based mostly on gravitational laws, you can, with very high credence, warrant a belief that as viewed from Earth, the

Moon will occlude the Sun at a certain future time. When "run backwards", the same models can be used to calculate the times of a past eclipses. And, on a more mundane level, as I left my car in the earlier example I could properly believe that it would be in the same place ten minutes on.

Indeed, the extent to which the past is physically fixed and the future is open is a metaphysical property with no direct epistemic application. Just because a specific event definitely happened at a specific time does not mean there is any way of knowing that it did. One can imagine a universe with space-time similar to ours but drastically varying laws, in which the total state at one time might reveal little about past states, however "fixed". Considering how the bulk of our beliefs are warranted, it is clear that the asymmetric bias towards beliefs about the past comes instead from the human capacities of perception, memory, and communication. We know more about the past because at times in the past we perceived what was happening, and are now able to recall, talk about, and extrapolate from what was perceived at different times to make further conclusions about past events. That is, we know more about the past mostly in virtue of having "been there".

Beyond this general point, I want to emphasize an important property of *reasoning*, which is that most of it is neutral with respect to past and future. Suppose, for example, that Obadiah has warranted beliefs that the knife *will* dip into the peanut butter jar, and that it *will* then move to one of the slices of bread and spread the peanut butter on it, and that the other slice *will* then be put on top of the first. The warrants could be in virtue of Agatha promising, or Obadiah could be an expert in the engineering and programming of a particular kind of robot that takes Agatha's place. Regardless of what kind of warrant he has for each of these beliefs, he can use them to justify a belief that there is about to be a peanut butter sandwich on the table. His further reasoning may not be *identical* to the retrospective case. He could, for example, have the beliefs I describe without knowing which slice will be which. But if knowledge of a sequence of past events is (together with background knowledge) sufficient to warrant some belief about the past, then knowledge of a similar sequence of future events might well warrant a similar belief about the future. Warranted belief about the future is limited more by a lack

of empirical data than by the ability to grasp the implications of data.

5.6 Epistemic Consistency

Given that the standard for high credence falls well below certainty, one could, absent further epistemic norms, have two mutually incompatible beliefs that are both warranted. Suppose that you see your cat napping in the hamper almost every morning, and believe that your cat is napping there this morning in virtue of that evidence. Then you see him walk by your desk. In virtue of this new sighting, together with some rudimentary metaphysics, you should not only believe that the cat is near your desk, but also revise your belief that it is in the hamper. This one sighting need not counter the general weight of past evidence; you might still reasonably believe your cat will be in the hamper tomorrow morning. But it certainly outweighs that evidence *right now*.

This sort of epistemic incompatibility is called "inconsistency". Not all forms of inconsistency are considered epistemically improper. A much-discussed acceptable example is believing that some of your beliefs are false (without knowing which). Still, there is general agreement that people have the capacity to avoid epistemic consistency within certain cognitive limits. That we can often *resolve* inconsistency, instead of just suspending judgment on the matter, is largely in virtue of the warrant requirement, and more narrowly relations of justification. The "record" of justification likely plays a role in identifying inconsistent beliefs, but is certainly what allows the case for now-questionable beliefs to be re-weighed in light of new evidence.

CHAPTER 6

Aims, Traims, and Technical Entitlement to Basic Actions

The first "stake in the ground" of the theory of this and the next chapter is that Agatha and Obadiah use the same pattern of reasoning to justify confidence in their respective attitudes. In Obadiah's case, those chains trace back to a combination of recent perceptual beliefs, such as that the knife dipped into the jar, and background beliefs formed at an earlier time. Just going by their matching descriptions, his perceptual beliefs correspond to Agatha's intentions to perform certain movements, such as dipping the knife into the jar. Given that Agatha in some sense reasons *from* those intentions, as Obadiah reasons from his beliefs, it is sensible to establish a theory of those first, and then consider her technical justification in light of its "premises".

In the action theory literature, intentional actions that can be described in terms of movements of the agent's body are often called "basic actions". Much of the nature of basic action remains somewhat mysterious, in part because of how much of it lies beyond our reasoning. However, one widely-held theory of basic action is roughly this: Humans have the capacity to move parts of their bodies in certain ways. Barring unusual circumstances, such as being tied up or anesthetized, we can rightly intend to perform such movements with (what I call) high technical confidence[1] (Davidson 1971, p. 60). By arranging such motions in sequence, often aided by perception of our immediate surroundings, we can accomplish ends such as walking or throwing, which can in turn be means to ends further removed from movement, such as going to the store or scoring a point.

1. Putting the point this way invites easy counter-examples, such as wiggling one's ears or moving an individual toe. However, such cases have not received much theoretical attention, perhaps because it is difficult to construct a plausible end for which these action would be a means.

This view could easily be adapted to support the theory of the next chapter. Indeed, if that were not possible the overall approach would falter, because we clearly *can* move our bodies in certain ways with confidence when we want to. Whether we *typically* choose or arrange such movements into sequences in that way is another question. If we do, we are not usually conscious of doing so. And if we arrive at those sequences in virtue of reasoning, it is of a different, less accessible kind than epistemic reasoning.

This "minimal" theory of basic action is analogous to certain theories of perception that have fallen out of favor. The theory that people experience only shape, color, and depth and attribute properties like category (such as *animal*) or identity (such as *Bob*) in virtue of reasoning was once dominant in analytic philosophy. It is much less so now, in part because it does not conform to the subjective experience of arriving at those conclusions, and also because some experiments, including time trials and fMRI studies, suggest that these processes are more continuous with our lower-level perceptual capacities than with our capacity for reason.

In virtue of related "behavioral" evidence, in this chapter I consider the problem of basic action in analogy to contemporary theories of perception and perceptual belief. My approach is to ask how a person is able to intend to perform a basic action with appropriate confidence, including high confidence. Parts of the answer are specific to the (reconsidered) domain of basic action, while others relate to confident action generally, and overlap with the topic of the next chapter. I start with the technical counterpart of a perceptual belief, which I call a *basic aim*. I then consider how best to distinguish basic and non-basic action, and the nature of our entitlement to appropriate technical confidence to an intention to perform a basic action.

6.1 The Attitude of "Aim"

The second "stake" is that technical confidence is the counterpart of credence. In the model of the last chapter, belief is a distinctly high-credence epistemic propositional attitude. Its counterpart would therefore be a propositional attitude of distinctly high technical confidence. To call this attitude "intention" would beg some questions. First, it would take Bratman's side

in the debate I discuss in Section 2.3.3. Second, to be the strict counterpart of belief, the attitude could potentially apply to *any* proposition, however intractable, physically impossible, or self-contradictory. My intuitions about whether one can *intend* to produce a round square are fuzzy[2], and nothing important rests on the issue. I will therefore use the term "aim" and its cognates as the name of this attitude[3]. As belief is the attitude towards a proposition that it is true, aim is the attitude towards a proposition that one will make it true, absent a change of mind. That is, it combines a revisable commitment to realizing the proposition, with high confidence that it is within one's power to do so.

As I just suggested, an aim can have any propositional content. You can aim that George Washington be Henry VIII, or that colorless green ideas sleep furiously[4]. However, just because you can have such an aim does not mean you can have it *appropriately*. As with analogous beliefs, such aims could only result from some cognitive mistake or impairment. And as impossible beliefs can be treated as a special case of beliefs that violate certain epistemic norms, impossible aims can be treated as a special case of beliefs that violate technical norms: such contents are inherently inappropriate because appropriate high credence or technical confidence in them is impossible.

The attitude of aim diverges from most theories of intentional representation in important ways, most obvious being the sort of propositional contents that *can* be appropriate. The first stake in the theory is that Obadiah and Agatha use the same reasoning to arrive at high credence and confidence, respectively. Obadiah's belief is that *there is a peanut butter sandwich on the table*. Agatha's aim is therefore that *there will be a peanut butter sandwich on the table*, rather than that *[she] will make a peanut butter sandwich*, or any other content that refers to herself or some action.

One can have such aims, but they would correspond to different intentions that could

2. See Section 7.6.

3. Although the ordinary meaning of "aim" is close to how I use it in this theory, it should be understood as standing in for the properties and requirements I ascribe to it, rather than as expressing its ordinary meaning.

4. Assuming that you can have beliefs with these contents.

require different means, or different types of planning. Following the example of Korsgaard (2009, pp. 210–1), I might revise an intention to write a better book on Kant after someone else writes one, because the reasons I had are now addressed. Or I might instead intend that *I* write a better book on Kant. If I were to revise the latter intention because someone else writes one, it would not really be because my reasons were addressed. Although either intention could be realized by the same means, the circumstances in which I would drop or revise each are different.

The general form of potentially appropriate aims therefore corresponds most closely to that of potentially appropriate desires in a belief-desire pair theory. An aim is not just a desire, because it involves a revisable commitment to make its content true. But that commitment is a property of the attitude itself, rather than being (necessarily) encoded in the content of the aim. Unlike desire, aim asserts high technical confidence; to aim that *P* is to take yourself as making *P* the case. The confidence is conditional on your continuing to aim, but again, neither the confidence nor that condition is encoded in the propositional content. Technical confidence is, like credence, a level or degree, and impinges least on consciousness when the degree is high enough to regard the (future-directed) content as true. And its revocable nature —it's "up to you-ness"—has no distinct content *or* degree. It is your sense that an aim is voluntary—that unlike with a belief, it can be freely changed or dropped.

Any talk of "taking yourself" and "your sense that" raises difficult questions of phenomenology. I say what I can about what it is like to aim in Section 7.4, after more of the theory is in place. Until then, the difference between aim and belief that is most important to keep in mind is that the former is experienced as voluntary. Setting aside certain metacognitive capacities, an aim might be best thought of as an assessment of *voluntary truth*[5].

5. This is how I understand the meaning of some terms related to the subject of *accuracy*: I take "accurate" to be neutral with respect to direction of fit. "Veridical" describes the accuracy of broadly epistemic representations, while "realized" describes the accuracy of broadly technical representations without having any further implications, causal or otherwise. Each means "accurate", but gives a nod to its corresponding direction of fit. A veridical belief is "true", but truth in isolation is not a type of veridicality. Assuming there is also *denial*—the attitude towards a proposition that it is false—a veridical denial is "false", but it would be strange to say that falsity in isolation is a type of veridicality. Realized aims and traims are, correspondingly, "true". Propositional attitudes that are not accuracy-dependent are neither veridical nor realized, although each is true, false, or suffers

6.2 The Attitude of "Traim"

This relationship between belief and aim, together with my choice in Section 5.4 of a distinct lower-confidence epistemic propositional attitude of suspicion, call for a corresponding lower-confidence technical epistemic attitude. I will call this attitude "traim"[6]. Traim associates some degree of technical confidence with some proposition. That degree might be specific, as when anticipating even odds that your coin toss will come up heads, or vague, as when anticipating a noise when you blow into a trumpet for the first time. Broadly speaking, any content that could be appropriate to an aim could be appropriate to a traim. There may be additional contents that can only be appropriate to a traim, corresponding to outcomes for which high confidence can never be appropriate[7].

---

from vagueness.

6. Although we frequently speak of "trying", which suggests lower confidence, there does not seem to be any conventional term for a distinctly lower-confidence intentional attitude. The lack of such a term may be one source of resistance to Bratman's argument that the concept *intention* only applies in (what I call) high-confidence cases—the thought being that there must be *some* easy way to refer to that kind of mental state. "Intending to try" is at least close, but still questionable; we typically use that phrase *before* an attempt, and it is not clear whether someone currently trying still intends to try. Anyway, given that my use of "aim" is definitional rather than ordinary, the fact that "traim" (as a portmanteau of "try" and "aim") is not even a word may be preferable. To be clear, its suggestion of "try" is only an indication of how we typically describe such cases. No theoretical position on the question of whether we also try in high confidence cases is implied.

7. Richard Holton discusses a conception very close to this one, also in relation to lower-credence (or "partial") belief (Holton 2008, p. 41):

> There are two ways to understand the idea of partial intention. We might say that an agent has a partial intention whenever they have merely a partial belief in its success. Or we might say that it is essential to partial intentions that they be only a proper part of an overall plan, that is that they be accompanied by alternative partial intentions to achieve the same end. As we saw, in agents' partial beliefs the analogues of these two different features come together: if I have a partial belief in p I will automatically have an accompanying partial belief in its complement, not-p. With intentions the result does not come automatically. If I have only a partial belief that I will achieve my end by succeeding in a certain intention, it does not automatically follow that I have an alternative intention designed to achieve the same end. I might have no back-up plan: I might simply have a partial belief that I will achieve the end at all.
>
> So we need to make a choice of how to define partial intentions, in terms of partial belief in success or of the presence of alternative intentions to the same end. I take the second path … After all, it does seem that if something is partial there should be other parts that make up the whole; that is the idea the definition captures.

Traim closely corresponds to Holton's "first path".

6.3 The Form of a Basic Aim

With the attitudes of aim and traim in hand, I can move on to the problem of confident basic action. A significant part of that problem, as I discuss in Section 6.4.2, is to establish what actions are basic. However, the first subject I need to discuss applies to any theory that conforms to the other common understanding of basic action as whatever one (intentionally) does not in virtue of (intentionally) doing something else. That is, as an action for which one takes no further means (Roth 2000, p. 12).

For the moment, assume that Agatha's spreading the peanut butter on the slice of bread is such an action. Her intention to do so corresponds to Obadiah's perceptual belief that it was spread, but here we need to take care about the content, or perhaps the context, of his belief. For reasons I discuss in Section 5.2.1, in order to function as part of his justification, it cannot simply be an isolated belief about a particular spreading. Obadiah must also be aware of the relation in time or sequence between the spreading and the other events of sandwich-making. On the assumption that Agatha uses the same pattern of reasoning, her representation of multi-step intentions must somehow encode corresponding relations. A general way of addressing this problem is to include a *condition* in the content of the intention, to encode the circumstances in which acting on it is appropriate.

Almost all of our basic intentions are conditional in the broad sense that we intend to act on them when some condition is met, which may be at a certain time, on the successful realization of an earlier condition, or when other circumstances (however defined) become favorable[8]. The same is not obviously true of an intention for which one has further means. If Agatha has decided when to dip, spread, and place, there is no place for a further decision about when to make a sandwich[9]. Even when an end is not yet realized after taking one's last

8. The term "conditional intention" has typically been used in the action theory literature in a narrower sense, to pick out those intentions that have a condition that may not occur, or that the agent is not confident will occur.

9. A partial plan might constitute an exception to this rule, given that when an intention to make a sandwich is not yet broken down into steps, it may need a condition; it would be unusual to intend to make a sandwich without having any conception of when to make it. On the other hand, it might not constitute an exception, given that a partial plan is one for which there are not *yet* any further means.

means, as in the much-discussed case of an intentional shooting now leading to an intentional killing later (Thomson 1971), there is still no role for any further condition. It is intentions without further means that *need* conditions.

There are many different patterns of condition that could time Agatha's intention to spread after her intention to dip the knife and before her intention to place the slice. For example, she could intend to perform the actions at 10:01, 10:00, and 10:02 respectively. However, most of our multi-step plans depend less on specific or even relative times than on *sequence*. A more appropriate condition for Agatha's spreading would therefore be that she has just dipped. We might therefore consider the content of her basic intention to be "[On successfully dipping the knife into the jar] I spread the peanut butter on the slice", with brackets to mark the open question of how sequence is encoded in mental states. I come back to this example in Section 6.4.4.

Another reason for bracketing the conditional content is that technical confidence in a basic aim should not depend on whether the condition will occur. In assessing it, one should the assume that condition will come about, and then judge one's likelihood of success given those circumstances. Technical confidence in a basic aim therefore does not depend on whether the condition is specific (like Agatha's) or vague, as in an aim to raise your hand *this afternoon* or *sometime*. Similarly, a basic traim should associate a degree of confidence in the likely success of the basic action on the assumption its condition is met. This does not mean that one can never be confident that a condition will be met, and therefore that one is highly likely to act. One could have such confidence in virtue of non-basic aims, which are the subject of the next chapter[10].

10. This view entails that an end-aim to raise my arm this afternoon is non-basic. The confidence in that end would be partly virtue of a means-aim that [this afternoon] I raise my arm, and partly in virtue of my quite reasonable belief that the condition *this afternoon* will be met. Although this division of labor may not accurately reflect how we reason in such cases, as a simplification it is handy, and (I think) mostly harmless. The theory could alternatively be extended to allow justificatory support for the condition of an otherwise entitled basic aim or traim, collapsing the two levels together.

6.4 Technical Entitlement to Basic Aims and Traims

Intention is only suited to its role in human action if assessments of technical confidence are generally appropriate—if we can generally bring about our aims, and the outcome of our traims generally correlates with the confidence we assign to them. With respect to at least some basic actions, such as raising an arm, high confidence seems to "come naturally". Absent restraints or paralysis, confidence that one can raise one's arm is not in virtue of conscious steps of reasoning or calculation. Asked to explain how she knows she can raise her arm, a child might say "That's just something arms can do."

As I discuss in Section 5.3, when people seem to arrive at beliefs of some type without reasoning, and cognitive scientists and philosophers can explain how that type of belief is generally true, it is appropriate to talk of an entitlement to that type of belief. The analogy between aim and belief can suggests the possibility of a *technical entitlement* to certain aims, such as the aim to raise one's arm. However, the nature of technical entitlement would need to fundamentally differ from, for example, perceptual entitlement. In Section 5.5, I observe that much of our reasoning is neutral with respect to past and future. This is not true of perceptual entitlement—a perceptual belief can only be derived from present experience or a memory of past experience. A technical entitlement to raising one's arm must instead come at least a little before the raising. If there is technical entitlement to basic action, it is prospective.

6.4.1 Prospective Judgment of Potential Bodily Movement

According to the more standard accounts of basic action, what we can "do basically" is confidently perform certain bodily movements. A theory of technical entitlement corresponding to this model would in some ways be simple: If the combination of brain and physiology in typical humans makes such movements reliable under normal circumstances—no compromising physical damage, sufficient energy from food, and so forth—a "default" assessment of high confidence would be reliable, and therefore appropriate. That reliability could then be explained in terms of the contribution to fitness of the capacity to first decide on movements,

and then carry them out.

However, this summary fails to address a difficulty introduced by the prospective and voluntary nature of intention. Because a perceptual representation comes before any belief is derived from it, the domain of potential perceptual beliefs is largely determined by the domain of perceptual representation. If I believe that I have smelled something pink, for example, we can account for what has gone wrong (or at least unusually) in my arriving at this conclusion in terms of the different modes of perceptual representation, and what they can represent. Because intentions are instead formed before we act, the domain of appropriate basic aims must be determined in some other way. This is the question of how one is aware that one can "bodily move" an arm but not a coffee cup, or bend at the knee but not at the thigh. I am not going to answer this question, not least because I can't, but I will briefly discuss some aspects of the problem that are relevant to technical entitlement.

One thought is that awareness (broadly speaking) of potential bodily movements is in virtue of *feeling* the body part, of having associated sensations of touch or pain related to it. It is this intuition that has likely prompted certain thought experiments about anesthetized body parts. *Feeling* does reliably differentiate between your arm and a coffee cup, but it is less clear how it differentiates between a potential knee and thigh bend. A better candidate for that job would be our capacity for proprioception—for sensing the relative arrangement of one's body in space. This capacity is rather different than the "classic" perceptual systems such as vision, hearing, smell, taste and touch, and features prominently in Anscombe (1957, §8) for that reason[11].

The problem with positing that directed movement is dependent on feeling or proprioception is that some people who lack both of those sensory capacities can still manage to perform basic actions, albeit with significant impairment. The problem has been long recognized, as Hornsby (1980, p. 40–4) notes. Feeling is also not sufficient for even a sense of the possibility of directed movement. A sufferer of "alien hand syndrome" knows the position of and

11. One way to describe the difference is in terms of there being no mode-specific *qualia* of proprioception. The way in which we are aware of the position of a limb is less like seeing or feeling where it is than like suspicions gained through "blindsight", or beliefs gained through its (theoretical) cousin "superblindsight" (Block 1995).

experiences sensations from an arm that she nevertheless does not take to be under her control (Doody and Jankovic 1992)[12]. Gallagher (2006) offers further examples, and considers the problem in virtue of the complex relationship between what Gallagher calls "body image" and "body schema", the latter being the system that eventually produces movement.

Whatever the factors that contribute to our awareness of a potential movement turn out to be, the model of perceptual entitlement suggests that one could be appropriately confident when the relevant movement is impossible, and even when the relevant body part is missing. Entitlement to a perceptual belief is not automatically nullified in a case of inaccurate perceptual representation. If you visually hallucinate, the beliefs you form about what you hallucinate can still be warranted, and this can be true even if you have no eyes at the time. Similarly, a person with a paralyzed limb, or someone suffering from a "phantom limb", could be technically entitled to an aim to move it or "it". Such an entitlement would not be absolute. If someone is aware she is missing her eyes, or the content of her visual experience strongly suggests she is hallucinating, she should work against the urge to form perceptual beliefs. Similarly, someone who knows his leg is paralyzed or missing should avoid forming a basic aim to move it. In either case, Burge's conception of a "default" entitlement, which could be overridden by other sources of information, seems appropriate.

6.4.2 (Lack of) Reasoning as a Criterion of Basic Action

Whatever the mental preconditions for directed movement are, there is a separate question of what kinds of actions are basic. One standard I have already alluded to is: bodily movements, so described. On this view, my reaching out to grasp a glass is not basic, or at least not basic "under that description". My basic actions would instead by the individual movements of my arm, hand, and fingers, however those movements should be individuated.

While there is nothing inconsistent in drawing the line this way, we should not mistake it for the boundary where reasoning starts. Luthra (2016, §3) argues against the view that we

12. It is in virtue of this subjective sense of "ownership" of the body part to be moved that I phrase the content of basic aims in the first person.

construct patterns of movement—to, for example, grasp an object—in virtue of reasoning[13]. There is evidence that conscious thought about the individual steps of some complex manual tasks often *lowers* one's chances of success rather than raising them (Flegal and Anderson 2008). And if I were to ask after you throw a ball just how you moved your leg, you might have to try again, or imagine yourself doing so, to answer accurately.

Experiments on participants undergoing brain surgery have also demonstrated that stimulating some areas of the brain can trigger complex coordinated movements like reaching and grasping, which suggests that those movements are already at least partly coordinated without our conscious influence (Fried et al. 1991). It has been argued that these "action plans" are what underlie the behavior of an "alien hand" (McBride et al. 2013). And the anatomical structure of the brain suggests that "hand-eye coordination" in movements like reaching is partly in virtue of perceptual information beyond that contributed through conscious awareness and participation: The primary visual cortex is connected by a wide "ventral stream" to areas of the brain associated with memory and recognition, but also by a wide "dorsal stream" to areas associated with spatial representation and guidance of movement.

By the no-further-means standard, if grabbing a glass is not a basic action, then we must do it in virtue of (intentionally) doing other things. Some candidates for those means are not plausible. By raising my arm I certainly cause certain nerves to fire, but I do not intentionally fire them. With that in mind, we might instead ground the distinction in the applicability of the concept *intentionally*. Unfortunately, the extension of that concept is also a matter of dispute. Perhaps the individual movements of your fingers as you reach for the glass are intentional, or perhaps they belong to a looser class of things done "with the intention" to grasp it[14].

One factor that weighs in favor of a reasoning standard is that it corresponds to how we think about analogous epistemic cases. Agatha's basic actions in making the sandwich roughly correspond to Obadiah's perceptual beliefs, and the distinction between perceptual entitlement

13. Luthra's primary concern in the paper is with certain conceptions of rationalism, rather than with what constitutes basic action, but the relevant considerations are quite similar. Note that he also uses the term "technical", but with a different, more restricted meaning.

14. See Anscombe (1957, §19).

and justification is one of reasoning. If you so wished, you could count four dots using the same process as you would to count a hundred of them, but you can also simply *see* four dots in a way that you cannot see a hundred of them *as* a hundred. In counting, you arrive at a sort of justification, in that you can give reasons for your belief that there are a certain number of dots. The warrant for a belief derived from simply seeing four dots would, however, be an entitlement. In a similar vein, you might (with a bit of practice) reach and grasp an object in front of you by thinking through the movements of each joint of your fingers, hand, and arm. Even so, that is not normally what you do when you reach.

It may therefore appear that there are two workable standards for what constitutes a basic action, one based on a distinction of reasoning and one based on a distinction of movement. However, when we consider each option in relation to the question of appropriate confidence, the movement standard faces further difficulties.

6.4.3 Entitlement to a Degree of Technical Confidence

One factor that is supposed to weigh in favor of a movement standard is that, barring certain unusual movements such as ear-wiggling or moving an individual, smaller toe, we can generally perform such movements with high confidence. If throwing a ball to someone is a basic action, then it is one that you can and often should do *without* high confidence. It might be argued that you still throw the ball with high confidence in some sense, just not the one having to do with where you want the ball to go. However, not every type of basic action has this kind of limited success—many failed skateboarding tricks are just crashes.

Regardless, "rescuing" the high confidence of throwing in this way is a cold comfort if it leaves you with little basis for an accurate assessment of your confidence in throwing *to*. You could reason about the latter by taking the distance and your past accuracy into account, but we generally take such steps only when we already have some doubt, and it is rare to arrive *back* at high confidence through reasoning. If you were instead throwing the ball to someone quite close by, you might not give the confidence in your overall success a second thought. The *initial* doubt when throwing far versus throwing near suggests that our capacity for basic

action also incorporates a faculty for assessing prospective confidence not arrived at through reason.

That one does not reason to a degree of confidence does not entail that one does *nothing* to arrive at it. To assess confidence in throwing, for example, you might imagine doing so. There is neuroscientific evidence that imagining the performance of a basic action activates the same areas of the brain as an actual performance, which is consistent with imagination providing a kind of access to the relevant capacities (Jeannerod 2001). To the extent that reliability in action is advantageous, this cognitive arrangement makes sense from an evolutionary standpoint. Some creature very different from us might have great dexterity through reasoning alone, but we clearly do not, as evidenced by how we gain some skills through repeated attempts, and our clumsiness when starting out. We evolved from animals with much more primitive decision-making capacities, but that no doubt still had various strategies for obtaining food or eluding predators that were neither fully nor equally reliable[15]. Having a more reliable skill without prospective awareness of its relative reliability is of limited value, and for an animal that is less "clever" than us, its reliance on reliable skills is all the more important to its survival. A bias in decision-making towards methods of greater skill requires some representation of them *as* greater.

A faculty for assessing confidence in basic actions is therefore consistent with what we know about our evolutionary history. In contrast with our entitlement to perceptual beliefs and suspicions, technical entitlement is prospective. Such a capacity would account for a technical entitlement to aims or traims that lie beyond (or below) the reasoning boundary, corresponding to the perceptual entitlement of Section 5.3.1. However, the argument in the case of action is if anything stronger, given how many of our high-confidence skills are gained through lower-confidence stages of learning. There is some evidence that we can also gain new perceptual "skills", such as the ability to "sex" chickens (Horsey 2002), but these do not

15. I think this counts as common knowledge, but for reference there are good examples of decision-making, learning, (apparent) confidence, and dexterity on the part of squirrels in the BBC documentary "Daylight Robbery 2" of 1991, as presented by Dr. Jessica Holm.

seem as central to human life[16].

Once we account in this way for appropriate confidence in *sub-rational* aims and traims, it becomes harder to justify the movement standard. There may be an important distinction between movements and sub-rational actions, but it is not that one can always have confidence in the former. You should not be confident that you can raise your hand, for example, if it is right below a shelf. *Without* a distinct category of sub-rational actions, a theorist can explain that lack of confidence as they do any lower-confidence case—probably in terms of some spatial reasoning. *With* that category, your confidence that you can raise your hand is better explained in part by your assessing that its path is free in the same way you judge your chances of hitting a target. In short, confidence in any movement depends to some extent on one's environment, and the reasoning line is the only relevant *prospective* distinction. I therefore consider the category of *basic action* to be equivalent to *sub-rational action*.

6.4.4 Examples

Agatha's aim in Section 6.3 is to spread the peanut butter on the slice [on successfully dipping the knife into the jar]. The theory of technical entitlement accounts for her appropriately high technical confidence in the following way. At the most basic level, Agatha is entitled to movements of her arm, hands, and fingers in virtue of a sense or map of those body parts that at present is only partially understood. She is entitled to *this* spreading (in the circumstances

16. In his dissertation, Luthra argues for an analogous "practical entitlement" (Luthra 2013, ch. 7). However, it takes the form of a general entitlement to our physical capacities, on the model of Burge's "default" entitlement, rather than an entitlement relative to a specific intention to carry out a basic action. On this view, the source of my lack of confidence in juggling ability (given that I've never learned how) would not be my motor capacities, but my knowledge (in my case from testimony) that the ability to juggle generally requires practice (p. 193).

I do not think his conception is incompatible with our shared view that some complex actions do not depend on reasoning, but I doubt it is compatible with confidence in intention being generally reliable. Even assuming we have a great deal of relevant knowledge of the limits of our capacities, and what tasks must generally be learned, the things we do are varied in so many different ways that I doubt we could reason to accurate probabilities of success. If the "default" were a *lack* of confidence, it might be plausible that we generally reason to higher confidence (when it is ultimately called for) in cases where it is important to us. But if the default is high confidence, why would you *start* to consider the question of confidence when you, for example, jump in this direction by that distance? I doubt that all our basic actions fit into categories that are sufficiently tidy to indicate, prior to any explicit reasoning, that such reasoning is needed. (This point corresponds to my argument for lower-credence entitlements in Section 5.4.1.)

she anticipates) in virtue of prospective access to her "motor" capacities, which may be unconscious or in virtue of her imagining performing the action under those circumstances[17]. These entitlements together make her aim, with its high confidence, meet a normative requirement of the attitude for *technical warrant*. If she considered a different content, such as throwing the knife into the jar from a distance, she might only be entitled to a lower degree of technical confidence. In that case, her aiming to do so would be inappropriate, but she could technically warrant a traim with that content, and accordingly describe herself as trying to throw the knife into the jar.

A generally aware person in a setting with no obstructions is entitled to an aim to raise her hand in virtue of the same access. If that person's arm is held or tied down, she is not entitled to that aim given that in the simplest imagined scenario her arm would be obstructed on that path by the other hand or rope keeping it down. However, she could still traim to raise it in those circumstances[18]. Further reasoning could rightfully restore her high confidence ("That's not a knot, but just a loop in the rope.") or fail to ("Perhaps it just *feels* like my hand is being held down."), but in either case we enter a different realm of assessment. Someone without any feeling or sense of proprioception of her arm, such as Landry's patient discussed in Hornsby (1980, p. 40), might be entitled to aim to raise her arm if she is not aware that it is restrained[19].

17. There are numerous further issues I could try to unpack here. By including "the slice" in the content I suggest that a slice has already been picked to be spread on, which is unlikely to be true much before the spreading. So that part of the intention might instead be made some kind of reference to an earlier if trivial slice-picking intention. It is also unlikely that anyone has a peanut-butter spreading skill so much as a skill for spreading adhesive viscous substances, so including *peanut butter* in the content of a basic intention may be misleading.

But similar caveats apply on the epistemic side. Obadiah's belief is about a particular slice, although his only representation of that particularity may be in terms of the brief history of events he witnessed. And it is unclear whether he can have a purely perceptual belief about peanut butter being spread, or must instead make an inference based on what he saw and his background knowledge.

18. In Section 10.2.5 I touch on the relation between very low technical confidence and intentional action. Whatever that level is, it does not seem to characterize what traims are appropriate. Even if one cannot win the lottery intentionally, one can certainly try to win the lottery. It is sometimes argued that one must see an outcome as possible in order to pursue it (Adams 1995), but in my view that standard is wrong unless it is interpreted as metaphysical possibility, which amounts to the goal needing to be coherent. I can *try* to put my hand through solid steel—perhaps it will magically turn to goop just before. On my view, no matter how the person's arm is held in place, she can traim to raise it.

19. The more unusual aspect of the case of Landry's patient that he is not aware he failed to raise his arm when it was restrained—a demonstration of his lack of proprioception. This is not relevant to the issue of *prospective*

Similarly, if her arm is paralyzed but she is not aware of that, she can aim to raise it. If she is aware, metacognitive reasoning should defeat the entitlement, but she can still traim to raise it. The same are true of someone who is not aware that her arm has been amputated, or is aware, respectively.

6.5 "Map-like" Intentions

Before moving on from the subject of technical confidence, I need to discuss another asymmetry between the formal structure of perception and perceptual belief on the one hand, and basic aims and basic action on the other. One consequence of this asymmetry is that some aims are either not propositional or are but do not bear the same relation to further reasoning as do most aims.

In Sections 5.1.1 and 5.1.2 I discuss the distinction between perceptual and propositional representations and their respective standards of accuracy. Some aspects of perceptual representation are accurate by degree, while a proposition is (setting aside the problem of vagueness) either true or false. Perceptual representation is also sometimes said to be "map-like" —the kind of thing, speaking very roughly and generally, that encodes properties by spatial relation, as a map does. A perceptual belief that is properly derived from a perceptual representation may be said to encode some of the same information in a different "format", allowing it to contribute to further propositional reasoning. That is, when an observer needs to reason about some information contained in a perceptual "map", she can (either "directly", if she is still perceiving the information, or "indirectly", in virtue of a perceptual memory) form a perceptual belief that encodes it. Because the perceptual experience precedes the belief, the transition is always *from* map-like *to* truth-conditional.

Given that aims and traims precede acting, this relation will not hold in technical reasoning. The difference would not be of much importance if the success of all actions were also all-or-nothing, and therefore always a good match for truth-conditional representation. But this is

technical confidence.

not the case—some of our actions succeed by degree. Consider an archer aiming her shot at a target. If we ask her to describe her goal, she might might say that it is to hit the center of the target, but this is not quite right. A bullseye shot is rare, and the score for an attempt is not all-or-nothing; it depends on the distance of the arrow from the center of the target.

To make this more clear, suppose that the rules of a match were slightly altered so that a hit in the small, center bullseye of the target receives *no* points. Unless the archer is very good, she might approach her shot just as she would in a normal match, if she judges that the increased risk of hitting the bullseye is outweighed by the likely decrease in score of her aiming off-center. However, it would make little sense to describe her goal as "to hit the center of the target" under these rules.

Confronted with these objections, the archer might instead describe her goal as "to hit as close to the center as possible" or "as she can". These descriptions do seem more apropos, and I suspect they roughly match how we often conceptualize such actions. The action itself is a process, with the success of the process being a matter of degree. The content of the aim or traim may stand in for the process, but, being propositional, cannot really "line up" with its success. At the same time, our intention does not merely represent the process, as if one's degree of success were beside the point. Whether the archer aims or traims may depend on whether she can reliably hit *somewhere* on the target, and therefore receive some positive score[20]. Therefore, even if her aim or traim is propositional in form, its truth condition does not bear the usual relationship to her further reasoning.

Although I have introduced this problem theoretically, as a consequence of a transition from propositional to map-like representation, there is also intuitive support for the difference. Suppose that you are satisfied with the archer's description of her aim as "to hit as close to the center as I can (in this instance)"[21]. Then if she were to succeed by something other than the highest possible degree, consider the accuracy of this description. This is the question:

20. That is, success can be a matter of degree even when there is a sharp line between *some* success and total failure.

21. Or if you are not satisfied, consider some related case and description.

"*Did* the archer hit as close to the center as she could in this instance?" Suppose she scores worse than usual. On one hand, you might think "no, she could have done better, given that she usually does." On the other, you might think "Well, in this instance this is how she did, so maybe it was the best she could do."

What I find most striking thing about this question is not that the answer is unclear, although that is unusual for what is supposed to be an expression of intention. It is rather how irrelevant the question becomes after the attempt is made. It certainly has no obvious tie to what we might deem intentional or not. Just before the archer releases her bow, she might have daydreamed about her upcoming vacation. Although that might be a reason to answer my *question* negatively, we would not then judge the entire shot to be unintentional. Whatever the prior representation of the intention, after the attempt it is the degree of success that matters, and questions about that representation fall flat[22].

What does seem right about such intentions is that whenever success is a matter of degree, prospective propositional reasoning cannot make much prospective *use* of that degree. If some minimum success is important, then reasoning about it calls for representing that minimum as a truth condition. Degree of technical confidence and degree of success are not at all the same thing. Conventional propositional *and* probabilistic reasoning are grounded in all-or-nothing conditions. The propositional representation of a goal that succeeds by degree can therefore be largely harmless to reasoning, as long as it is not misinterpreted in subsequent reasoning.

6.6 Further Requirements of Aim and Traim

A normative requirement that aims and traims be technically warranted only addresses the *potential* for an agent to carry out some basic action. Agatha could consider the possibility of spreading peanut butter on a slice of bread, and assess she can do so reliably, while never

22. It has been suggested to me that this impression could just be a consequence of vagueness in the statement of the goal. I disagree with this interpretation for two reasons. One is that I just don't think it lines up with the intuition: I am familiar with what vagueness "feels like", and this "falling flat" is different. The other is that I see no reason there would be a systematic relationship between goals that are successful by degree and vagueness in all intentional descriptions of those goals.

deciding to or forming any other kind of commitment. I argue in Section 2.1 that we often intend confidently, in that we take a goal as something that will come about unless we change our minds. To serve as the basis of such intentions, requirements on the attitude of aim must also account for one's confidence that one will make an attempt. These requirements would account for the commitment involved in intending, and how that commitment typically leads to action. In the case of basic action, they must account for how one will generally come to act on a basic aim once its condition is met, absent a change of mind. The attitude of traim can be served by the same explanation, because the lower confidence of a traim is in the chances of success, not in the chances that one will make an attempt.

It is easiest to consider these requirements moving back in time from the point of acting.

6.6.1 The Will

Obadiah witnesses Agatha spread the peanut butter on the slice after dipping the knife in the jar. Setting aside whether her doing so was intentional, suppose that she takes the spreading to be the realization of her intention to spread, and Obadiah believes this also. How should he describe Agatha's mental activity as she acts?

Some theories of intentional action set aside the question of consciousness, and just characterize intention as a mental state that *causes* the associated behavior, with the caveat that it must cause it in the "right way". I do not believe that this approach is wrong, exactly, so much as that it is too vague to be useful. The same claim might be made about the role of belief in justification, but epistemologists, to their credit, generally try to be more specific about that relationship. One specific objection to this *mere Causalism*, in Velleman (1992) and elsewhere, is that to say an intention causes the action is to leave the agent herself out of the theory. This is not a definitive argument—one response is that the agent is included in virtue of the intention being hers—but I share a version of it.

As I have defined it, aim is an *attitude* towards a propositional content, as belief is an attitude. When such an attitude causes further mental states or processes in the "right way",

it is most often in virtue of a person's conscious awareness of it. Indeed, while it is difficult to precisely distinguish between mental "states" and "processes", we most often talk of states as being the objects or "parameters" of processes. An attitude is in this sense a mental state: something one can be aware of, which might affect subsequent mental activity. On this picture, the "right way" that Agatha's intention to spread causes the spreading is that Agatha is aware of the intention and spreads partly in virtue of that awareness. This arrangement raises two questions: Why or how does the intention generally prompt Agatha to act, and what is this "further step" on the way to behavior?

Suppose that Agatha is making the sandwich for Arthur, who has a deadly peanut allergy that she is aware of but he is not. She has decided to kill him in order to inherit his fortune, but after she dips the knife into the jar she has a crisis of conscience and is unable to bring herself to spread. If she still in some sense feels that she *should* go through with her plan, we might call her failure to spread a "weakness of will"[23]. The term "will" (along with its cousin "volition") is used with various different meanings in the action theory literature[24]. Regardless, in light of the concept of weakness of will, I will use the term "will" to refer to the link Agatha fulfills between aim and behavior when she is *not* "weak": In virtue of her awareness of her aim to spread, she spreads in virtue of an exercise of her will.

The will, or this aspect of it, is a difficult and interesting subject, but not one I will say much more about. Nothing about technical rationality seems to hinge on whether we have *free will*, for example. And although I am picking out the operation of the will partly in terms of consciousness, nothing much hinges on that either. A robot designed on the model of this thesis might have a routine that iterates through a list of "aims", picks out those whose

23. The topic of weakness of will is one of several aspects of action theory that have been most widely discussed in the context of an ethical project. Perhaps for this reason, examples of it in philosophical papers are usually constructed so that the reader judges that the agent should go through with her plan. This bias is sometimes so extreme as to leave the author's proposal unclear. Unless I missed them, Velleman (1992) discusses no conflicted actions portrayed as a right thing to do, or fully-committed actions portrayed as wrong. The reader must judge for herself whether those arrangements are just as possible, or are ruled out by some other factor. I favor a psychological approach to phenomena like weakness of will rather than an ethical one. Vogler (2009) explores similar issues and biases in much greater detail.

24. Or possibly with a complex, unified meaning that I do not yet understand.

"conditions" have been met, and chooses one to pass on to its motor sub-system. In terms of separation of responsibilities, we could call this subroutine the robot's "will".

Given some appropriate conception of will, however, whether one will generally act on one's aims and traims depends on one's will generally not being "weak". In other words, when an agent aims to φ when condition *e* is met, and is subsequently aware of both her aim and *that* condition *e* is met, she must typically "be willing" to act. Because weakness of will is usually the result of having practical reasons that (rightly or wrongly) count against acting[25], this amounts to a requirement that one's aims generally conform to one's practical but non-technical reasons. There may be unusual cases in which it is difficult or impossible to avoid weakness. You might make an apparently appropriate decision to act, and only later discover some reasons that you were unaware of, or that your reasons have different relative weights than you thought[26]. The point is that commitment to an aim or traim is not just a *stipulation*. If you can sometimes decide to act against what you judge to be your values and interests, the leeway in that direction is limited.

The existence of some version of a requirement that an intention accord with one's practical (as opposed to technical or "instrumental") reasons is, of course, already widely recognized. I raise the issue here because of the role it plays in reliable human action, but further details are beyond the scope of technical rationality.

6.6.2 Non-voluntary Recall

In order for Agatha to exercise her will and bring about spreading the peanut butter, she needs to be aware of her aim to spread it. Agatha's initial high confidence in her conditional intention

25. Other failures to act can occur at this point, but those would not be called "weakness of will".

26. I mean to distinguish here between appropriate aims that eventually suffer from weakness, and inappropriate ones. Someone who decides to act entirely against their reasons fails to meet practical norms that fall outside of technical rationality. Aims formed on that basis are inappropriate by that non-technical standard. Still, I think one can also reason correctly and still suffer from weakness of will. Consider a person who decides to dive off a higher board for the first time. She might have every reason to dive, and only realize she is more frightened than she expected when she first looks down at the water from that height. This is not necessarily a mistake of reasoning—she might just have lacked some relevant information.

to spread after successfully dipping the knife is only appropriate if it is highly likely to come into her awareness after she dips. We therefore cannot fully account for the commitment of intention without accounting for how she reliably comes to that awareness.

This way of "coming to mind" is not a general property of mental attitudes, and is not a property of belief. For Obadiah to believe that *P*, it does need to be highly likely that he becomes aware of the (potentially mistaken) truth of *P* when the question of "Whether *P*?" seems relevant to him. If he considers whether some peanut butter was spread on the slice, and cannot recall in the moment, it is at least questionable whether we should say that he believes that it was. He might at the same time be a capable of other forms of cognition that normally characterize belief. If he reflected on his belief that there is a peanut butter sandwich, his reflecting on *why* he believes might jog his memory, making him conscious of the spreading. Even so, belief is the attitude that a proposition is true, and if you are are not sure whether *P right now*, there is an important sense in which you do not believe that *P right now*.

Aside from the requirement that a belief come to mind when "Whether *P*?" is relevant, it could also come to mind at other times. Thinking about peanut butter later, Obadiah might recall its being spread. However, such associations are not reliable, and are not constitutive of belief. Obadiah might rightly think about Agatha or peanut butter sandwiches or even her dipping of the knife without also thinking about her having used it to spread. Even the recall of a justifying relation is not a requirement of belief *per se*: If Obadiah believes there is a sandwich, but on reflecting does not remember why he believes that, he should revise that belief. But his failure to recall the spreading in this circumstance does not necessarily call *that* belief into question. He may retain the latter, and have just lost the inter-propositional "glue" that constitutes justification.

In contrast, a desire that *D* need not come to mind when the question of "Whether *D*?" is relevant in order to *be* a desire. It does probably need to come to mind when asking oneself "Do I want it to be the case that *D*?". An aim that *Q* probably *does* need to come to mind when wondering "Whether *Q*?", but for different reasons than a belief does: whether *Q* will be the

case without one's intervention is important to planning[27].

Unlike with belief, however, if a basic aim were not highly likely to come to mind when its condition has been met, it would not be reliable. To act on her conditional intention, an agent must "put together" that intention with the fact that the condition has been realized. If humans routinely failed to do so, the high confidence of the attitude of aim would never be appropriate. This distinction between belief and aim is evident in what it means to "forget" an intention. "Forgetting" a dentist appointment is not usually like forgetting that Montevideo is the capital of Uruguay. If someone were to bring up the subject of dentists or teeth, the entirety of one's appointment-intention might immediately come to mind. Forgetting an intention is instead most often a failure of *non-voluntary recall*—of a constitutive requirement to be back in one's awareness when it is time to act.

How the recall requirement is met will depend on the condition. One might recall an intention to act at a specific time—which we should remember is probably not a scenario important to our evolutionary history—with a frequency that increases with one's sense of how much time is left. And one might recall an intention to act "sometime" when one has nothing else to do. Whatever the condition, aims generally to come to mind under different circumstances than beliefs. The relations of content, present cognition, and awareness that are constitutive of the mental states underlying action are therefore different from the corresponding relations that are constitutive of the mental states underlying knowledge. This difference is at the root of what makes each something like the inverse of the other[28].

Beyond the formal case for the non-voluntary recall of aims, there is cognitive evidence that having a conditional intention increases the likelihood of acting on a goal (Gollwitzer and

27. See Section 8.3.1.

28. The difference therefore poses a theoretical problem for traditional Cognivitism. Setiya could explain non-voluntary recall in virtue of his belief-intentions being "desire-like", although the requirement on aim seems stronger than that on desire. (I might really like donuts, and would enjoy one right now, but not experience a desire for the doughnut in front of me if there are more important things going on.) But if an intention is just a belief with a particular content, it is not clear why you would tend to recall an intention at the right time. To make this point explicitly, suppose you could voluntarily believe "I will consciously recall this very belief upon discovering that *e*." A Cognitivist must also account for the capacities that it will tend to make such beliefs true.

Sheeran 2006). Because the control group in Gollwitzer's experiment are participants with a corresponding partial plan, the contrast he finds can be interpreted as calling the reliability of partial plans into question, rather than supporting the reliability of aims in general[29]. What is definitely relevant, however, is the evidence that intending to act under a certain condition can shift cognitive resources towards detecting that that condition has been met (Gollwitzer 1993, pp. 160–5). And the periodic conscious recurrence of an intention can prompt one to consider or even research the status of its associated condition. Aiming to φ when e, therefore, does not just make it likely that one will recall the aim when one believes that e is the case, it also makes it more likely that one will *detect* that e is the case.

It is of course apparent that the capacity to maintain multiple simultaneous intentions is limited, varies by individual, and, as with our other cognitive capacities, decreases as we get older. And just as the increase in information a person may need has prompted a shift from reliance on memory to reference materials and search engines, the increase in the number and type of intentions a person needs to act on has prompted increasing reliance on reminder notes and calendar applications. Even so, forgetting to do what you most want and anticipate is rare. With intention, the two meanings of "excitement" converge. Anticipation of an exciting event can so stimulate cognition that it becomes difficult to sleep. There are many interesting facets to the requirement for non-voluntary recall that I will not explore in this thesis. What is important is that there is such a requirement, and it is very much a core aspect of technical rationality.

6.7 Basic Aims and "Types" of Intention

In Section 7.6, after I extend the theory to account for confidence in non-basic action, I discuss different ways the concept *intention* might be defined (in a loose sense) in terms of aim and traim and their technical requirements. There is one part of that larger question that I can ad-

29. The difference is relevant to technical rationality, but I suspect the general failure rates say more about the "behavior" of low-motivation intentions than about the reliability of deferred planning. (See Section 8.4.1).

dress now, however, as it seems to apply only or primarily to basic action. This is the question of whether there is a difference in kind between an intention to act *later* versus an intention to act *now*, and what that difference is. There are two pairs of terms most commonly associated with this possibility. Searle (1983, p. 87) describes a contrast between "prospective intention" and "intention in action". And many authors differentiate between "distal" and "proximal" intentions (Mele and Moser 1994, pp. 47,65). Setting aside the question of whether these two pairs of term have been used consistently, there seem to be two separable distinctions behind the various uses of both.

Morrison (2016, p. 36) notes the possibility of "action confidence". On the interpretation of Section 5.4.1, *action confidence* would bear the same relationship to technical confidence in an entitled aim or traim as *perceptual confidence* bears, in Morrison's theory, to the credence in a perceptual belief or suspicion. It would be the confidence you have when actually reaching for a handle, which would drop if the handle were to suddenly move away from you. I have argued that such actions count as basic, and that we do not *reason* to our prospective confidence that we can carry them out. To draw the line this way also suggests that there are elements of sub-rational experience that correspond to acting. These elements would not be *perceptual*, although they would often be closely intertwined with perceptual representation. Even "traditional" basic actions could have these elements.

In describing "intention in action", Searle says "Raising your arm, like seeing the table, characteristically consists of two components: the experience of raising your arm and the physical movement of the arm." (Searle 1983, p. 87) and "As far as Intentionality is concerned, the experience of acting just is the intention in action." (p. 91). Therefore Searle, at least, seems to be getting at something like the representations that Morrison associates with *action confidence*. These representations would not be propositional, would be accuracy-dependent, and would correspond to perceptual experience but have the opposite direction of fit[30]. Following

30. When you imagine yourself performing some basic action in order to judge prospective confidence, your experience would be partly in virtue of what amounts a hypothetical variant of that representation. Imagining of this kind can therefore be accuracy-dependent in a restricted sense, as hypothetical reasoning is accuracy-dependent in a related, restricted sense.

the terminology of Section 5.1.3, these could be called "broadly technical". The theory of basic aims and traims is committed to our having this type of representation, given that confident action ultimately depends on it. The relationship between belief and aim (and suspicion and traim) also implies that we have such representations.

Other authors who use these terms seem to distinguish two kinds of propositional (or otherwise conceptual) representation. "Proximal" and "distal" are more often used in this sense. The idea seems to be that there is a change of kind in intention on deciding or coming to be aware that it is time to act, or that one is acting. I do not fully understand the motivation for this view, unless it is supposed to capture, in a different way, what I characterize as coming to realize that the condition of an intention has come about. If that is basis of the distinction, the theory of this chapter already addresses that aspect of coming to act.

Otherwise, I see no more reason for a numerically different aim or traim than I do for a numerically different belief in an analogous case. Suppose that I believe there will be an eclipse at eleven tonight, and then eleven arrives and I believe that there is an eclipse *now*. While I have a different sort of evidence for my belief as the eclipse occurs, I do not see why I necessarily have a different belief. Similarly, when I intend to act in the future and then realize the time to act has come, I do not see why I necessarily have a different intention. If the evidence is supposed to be the descriptions I would offer at each time, that seems overly linguistic, and cuts against a widely accepted, although not universal, "tenseless" interpretation of propositional reasoning[31].

31. Pacherie (2008, §2) distinguishes three types of intention. Her "D-intentions" are closest to aims and traims, while I read her "P-intentions" as corresponding to the sort of broadly technical representations that interest Searle. Her third type of "M-intentions" has to do "fine-grained" motor control in isolation. In my view, the extent to which such fine motor control *cognition* involves distinct *representation* is open to question. The literature on this subject often seems to assume that the position of a limb, for example, *must* be represented in the process of moving it, rather than discussing particular signals and specifically what they represent. To the extent that there is true representation at this level, however, it can be treated as a sub-class of broadly technical representation.

6.8 Summary

So here, in summary, is a theory of intention for basic action constructed on the model of perceptual belief: Aim is the attitude corresponding to belief that one will (likely) make a proposition true, absent a change of mind. A basic aim has the content that a basic action will be performed under a certain condition. The normative requirement on a basic aim is that one have a technical entitlement to the reliable performance of the basic action when the condition holds. An aspect of that entitlement is that one has a sense or map of, and feels that one can control, the parts of one's body that will move in performing or attempting the action. A constitutive requirement on the attitude of aim is that an aim with conditional content will generally return to the agent's awareness once she has an epistemically warranted belief that the condition holds. Having a conditional aim makes it more likely that that one will seek out evidence relevant to whether the condition holds, and form the belief on getting that evidence. One can traim to perform a basic action in virtue of an entitlement to a lower degree of technical confidence, but otherwise the requirements for traiming are the same as those for aiming.

CHAPTER 7

Technical Justification

Although the previous chapter's theory of basic action addresses more than just bodily movement, it is still limited to actions whose descriptions directly imply some movement on the agent's part, such as reaching, grasping, throwing, or walking. In Anscombe's other famous example (Anscombe 1957, §23), it would account for the intention to pump but not the intention to poison[1] According to one widely-held theory, every "non-basic" action is *identical* to a basic one[2]. One difficulty this view faces is accounting for actions that require more than one step. Time delays, including the fact that the pumping and poisoning do not occur at the same time, pose another challenge.

Whether or not some combination of representation and identity can differentiate what happens intentionally from what does not[3], those factors alone cannot account for appropriate confidence. One might pump confidently while poisoning with high or lower confidence, depending on one's beliefs about the circumstances. In this chapter I develop a theory of appropriate confidence for non-basic action. Given that I draw the line of basic action where reasoning starts, it will be no surprise that the theory accounts for non-basic action in virtue of a kind of reasoning.

Our capacity to act non-basically with appropriate confidence depends in part on our capac-

1. The movements of the arms would likely be made "with the intention" of pumping, rather than being explicitly intended.

2. Anscombe chose her example in part because she considered it intuitive that the poisoning *is* the pumping, although she had no general theory of identity (Anscombe 1979, §6).

3. Representation factors in because not all events that result from a basic action could *could be* intentional *are* intentional. A poisoning may also be an orphaning, but whether it is an intentional orphaning depends on the pumper's plans.

ity for appropriately confident basic action—someone who does not pump confidently should not be confident of poisoning. However, the theory of this chapter does not otherwise depend on the theory of the last. It is also compatible with the traditional view, in which the connection between arm movement and *reaching for* is treated as a form of (perhaps unconscious) reasoning. As perceptual entitlement and epistemic justification are quite different ways of warranting a belief, technical entitlement and *technical justification* are quite different ways of warranting an aim.

The last chapter could also be characterized as complicating what was thought to be simple. Although it would be an exaggeration to say that this chapter does the opposite, some of the work is already done. The first "stake" of Chapter 4 is that ends and means are related by the same patterns of reasoning that relate justified and justifying beliefs. This is the theory in a nutshell. I will start to work through its implications, and describe it in more formal terms, by applying it to Agatha's reasoning.

7.1 Agatha's Reasoning

Recall how Obadiah and Agatha explain their respective belief and intention:

> Well, the knife dipped into jar, and got peanut butter on it. Then the knife moved to one of the slices of bread, and the peanut butter got spread on it. And then the peanut butter wound up between the two slices when the other one was put on top. And all a peanut butter sandwich is is two slices of bread with peanut butter between them.
>
> I will dip the knife into the peanut butter jar, so it will have peanut butter on it. Then I will move the knife to one of the slices of bread, and spread the peanut butter on it. Then I will pick up the other slice and put it on top of that one, so that the peanut butter is between them. And because all a peanut butter sandwich is is two slices of bread with peanut butter between them, I will have made one.

If the same reasoning underlies these descriptions, we can approach the question of Agatha's specific, complex mental state of non-basic intention by making gradual adjustments to Obadiah's justified belief, while keeping the structure of the reasoning intact.

The first adjustment is to change the tense. Obadiah can plausibly have a justified belief

that there *will be* a peanut butter sandwich if he were to have warranted beliefs that each of the individual steps he witnesses in the original example *will* happen. As I note in Section 5.5, justificatory reasoning is generally neutral with respect to past and future. Even though this change results in a different *overall* epistemic warrant—Obadiah's beliefs about the steps can no longer be perceptual—his "immediate" justification for their being a peanut butter sandwich is left intact[4].

The second adjustment is to replace the now-future tense beliefs about those steps with equivalent basic aims, each conditioned on the completion of the previous one[5]. After this adjustment, however, what had been justified beliefs are no longer epistemically warranted. Consider Obadiah's (adjusted) belief that there *will be* two slices of bread with peanut butter between them. In his unaltered reasoning, the counterpart to that belief is justified in part by perceptual beliefs. This one is instead "justified" by basic aims. A *belief* about what will happen cannot properly be justified by even a high-confidence intention to take some step. Speaking formally, a belief can only be justified by other beliefs and suspicions. Speaking intuitively, intentions are up to you in a way that beliefs are not.

For Agatha's reasoning to have the same "pattern" as Obadiah's, her attitude needs to have the same content as this now-unwarranted belief. What Agatha says about this part of her plan is "Then I will pick up the other slice and put it on top of that one, so that the peanut butter is between them." The arranging of peanut butter between two slices is therefore plausibly a thing that she intends to *do*, presumably with high confidence. At the level of basic action, the attitude that corresponds to high confidence intention is aim, and in the broad comparison between knowledge and intentional action, aim corresponds to belief. In this light, the third adjustment is to replace the now-unjustified beliefs with aims of the same content. Agatha's corresponding attitude can therefore be the *aim* that there will be two slices of bread with peanut butter between them.

---

4. I could tell a story about how these future-directed step beliefs are epistemically warranted if one were needed. Perhaps Obadiah has witnessed some machine perform them hundreds of times, and anticipates that the same events are about to happen.

5. The condition for the first step might be "when both Obadiah and I arrive at the table".

The final adjustment is to make the same substitution up the chain of justification, to those beliefs now "justified" by these non-basic aims. The last of these will be at the "head" of what was Obadiah's epistemic justification. In Agatha's reasoning, this becomes an aim that there be a peanut butter sandwich. As I discuss in Section 6.1, the difference between this propositional content and how Agatha would describe it with "I will make a peanut butter sandwich" is made up by the attitude of aim. In taking a content as your aim, you take yourself as responsible for its eventual realization[6].

In the last chapter I developed a technical equivalent of epistemic entitlement. Various threads of the last three chapters are converging on the claim that, with these last adjustments, Agatha's high technical confidence that there will be a peanut butter sandwich will instead be *technically justified*. The justification is in virtue of the same reasoning as Obadiah's, with a shift from perceptual beliefs to basic aims, and a content-preserving shift from further beliefs to non-basic aims. I can now make that claim explicitly, and defend it with answers to two questions.

The first question is whether her confidence that there will be a peanut butter sandwich is appropriate *apart* from the question of her intentional responsibility. Above the level of basic action and perception, Agatha and Obadiah support their respective aim and belief with the same reasoning. The truth of his perceptual beliefs, together with background beliefs that Agatha shares, secure the truth or high likelihood of his sandwich belief in virtue of his epistemic reasoning. If Agatha can secure the truth or high likelihood of her basic aims, the same reasoning should secure the truth or high likelihood of her sandwich aim, given that all her intermediate aims have the same (tenseless) content, and differ only in attitude. And she *has* secured the truth or high likelihood of her basic aims–they are warranted in virtue of her being technically entitled to them.

The second question is whether Agatha's reasoning also secures her own responsibility for making the sandwich, therefore distinguishing her high technical confidence from what

6. This is an over-intellectualized way of putting the point, however, in that the capacity to aim does not depend on grasping the concept *responsible*.

would (or should) be high credence. I have already noted why high credence is not appropriate: an aim cannot epistemically justify a belief. In Agatha's case, not only does an aim (partially) justify another aim, but *all* of her aims are either justified in part by aims, or are in virtue of a technical entitlement. Agatha's aim to make a peanut butter sandwich is justified in part by her aim that there be two slices of bread with peanut butter between them, which is itself justified in part by her aim to place one slice of bread on top of the other. These chains of technical reasoning therefore preserve not just the truth of her end, but her intentional contribution to it.

The *form* of Agatha's technical justification therefore exemplifies a different "division of labor" compared with the theories of Chapter 3. Most theories of intentional content require some ascription of responsibility to the agent, and perhaps a self-referential ascription of causality between the intention and the intended outcome. Cognitivist theories need such content because they suppose that intentions are beliefs. If belief is just the attitude towards a proposition that it is true, the features that distinguish intention-beliefs must lie in their contents. The present theory accounts for those conditions not with propositional content, but in virtue of the normative requirement for technical warrant.

An advocate of Searle's conditions-of-satisfaction approach would argue that those conditions must somehow be present in the attitude of aim itself. I find this position doubtful on evolutionary grounds. Certain metacognitive processes may require forming attitudes *about* attitudes that partly concern conditions of satisfaction. But I see no reason why the conditions of success of a cognitive sub-system are necessarily reflected in some aspect of *representation* in that sub-system. Regardless of what the attitude encodes, however, those conditions are not present in the content of aims, and we do not typically have to reason about them. Almost all technical justification would be at the "object level"[7].

7. See Sections 2.2 and 3.4.

7.2 "Preservation" of Voluntary Truth by Technical Justification

The form of Agatha's reasoning can be generalized into a theory of technical justification that supports common intuitions about the means-end relation, and can serve as a normative account of our reasoning about that relation. What it is to take something as a means is to comprehend that, *given how the world is*, taking those means will bring about the end. We are rarely sufficiently certain about our circumstances or capacities to be certain about that relation, but we still can reason in the following way: Suppose some proposition M is one you can generally bring about intentionally. And further suppose that M, together with other of your warranted beliefs, secures the truth or high likelihood of E. In that case, it is reasonable to have high confidence that bringing about M will bring about E.

The justification for Agatha's aim matches this general form: The technical justification of an aim E by a truth- or high-likelihood-preserving relation of one or more technically warranted aims $M_{1..m}$ and any number of epistemically warranted beliefs $B_{1..b}$ secures the role of $M_{1..m}$ as means by meeting three criteria. The first criterion ensures the truth or high likelihood of the end's realization given the realization of the means. This is in virtue of the *pattern* of reasoning, which must be such that if the aims $M_{1..m}$ were replaced by epistemically warranted beliefs of the same content, the same reasoning would epistemically justify a belief with the content of E. A "pattern of reasoning" in this sense is therefore the inter-propositional relation that *could* be used in an epistemic justification, without any presumption about the *attitudes taken* towards those propositions[8].

The second criterion ensures intentional responsibility—that the end's coming about depends on one taking one's means. Because $M_{1..m}$ are themselves technically warranted, they have the status of high-confidence intentions, which ultimately depend on technical entitlement to basic actions. The third criterion is that the end is distinguished from a foreseeable side-effect of one's means. By representing E *as* an aim justified by $M_{1..m}$, one takes E as

8. That is, the *contents* of the propositions are treated as fixed, while the *attitudes* towards them are treated as variable, or outside of the pattern.

something one will bring about by bringing about $M_{1..m}$ (absent a change of mind)[9].

This separation of truth and intentional responsibility into distinct aspects of reasoning may, however, already over-intellectualize the theory. I note in Section 6.1, the minimal interpretation of an aim is as a *voluntary truth*. Technical justification might therefore be best seen as "preserving" voluntary truth (or high likelihood) in the sense analogous to how epistemic justification is sometimes said to "preserve" what could be called "involuntary" truth (or high likelihood). Agatha considers both the technically justified end of making a sandwich and the justifying means of dipping, spreading, and placing as up to her, and she would violate a normative requirement if one of these aims were not technically warranted. If she were to to stop aiming to spread, she is required either to stop aiming to make a sandwich, or to justify the latter aim by some other means.

What connects her end and her means is not any additional content, but rather the inter-propositional "glue" of justification[10]. That relation also extends the voluntary nature of the means to the end. The normative requirement for technical warrant ensures that a justified aim is voluntarily true, without any need for agent-ascriptive, self-referential, or meta-cognitive content. Instead, one's basic aims are like additional premises that one can instantiate in the world, in the service of justifying a voluntary conclusion.

The relation of justification can also account for an agent's ongoing revision of her plans and degree of confidence as circumstances change or her attempts falter. Finding the peanut butter jar empty will prompt Agatha to revise her belief that it contains peanut butter. Because that belief is part of the technical justification of her aim to make a sandwich, that aim will

9. My focus in this thesis is on cases in which an agent does things "on her own", or in virtue of means that are considered intentional, even if she is not entirely responsible for them. I take it that you can intentionally vacation in Manchester partly in virtue of intentionally traveling there in a plane even when you are not its pilot. Some other cases of cooperation—and, for that matter, coercion—are not like this. If Agatha asks Obadiah to make a peanut butter sandwich, and he does so, the sandwich is *his* intentional accomplishment. It could also be hers, as when she has social authority to *order* him to make one, and she does so. But then it is hers only in a different and qualified sense. A complete theory of cooperation, including cases of *shared agency*, will therefore require a more complex specification of technical justification, in which the intentional actions of others either do not qualify as means, or are a special kind of means.

10. See Section 5.2.

now lack warrant, requiring of her some revision of content, confidence, or justification. This coming-to-awareness that what was a high-confident intention is no longer warranted is the complement of the non-voluntary recall of a basic aim. To intend with appropriate confidence requires that one be aware of one's intention to act at the appropriate time. But it also requires that one become aware when confidence is no longer appropriate, so that one can rationally revise plans and commitments[11].

7.2.1 Arriving at a Technical Justification

The direct application of epistemic patterns of reasoning to technical justification accounts for one cognitive norm in terms of another. In Section 1.2.2 I give the example of accounting for a consistency requirement in terms of requirements for closure and against contradiction. This new relationship is not a reduction of one norm to another, but a commonality of form between two norms, one of which is better understood and accepted. However, a proposal at the level of norms is at best a partial solution, given that someone can still raise this objection:

> I can accept that a technical justification in virtue of means-aims and beliefs could make high confidence appropriate. But I do not accept that our capacity to arrive at an epistemic justification accounts for a capacity to arrive at a technical justification. When we reason instrumentally we start with our end, which in your proposal is equivalent to an epistemic *conclusion*. And reasoning from a conclusion is exactly what you're *not* supposed to do.

Part of the basis of this objection is sound. Many of our beliefs, particularly those we do not arrive at through conscious reasoning, are likely produced by a process that *starts* with beliefs we already have. At best, such processes would need significant alterations to serve the needs of technical rationality.

We also, however, arrive at new beliefs in other ways, one of which starts with the question "Is it the case that *P*?" Suppose that Obadiah did not witness Agatha's behavior, and is wondering whether that thing on the table is a peanut butter sandwich. Given his warranted

11. Strictly speaking, the revision of means upon abandoning an end is not a requirement of *technical* rationality; one might still be in a position to carry out some or all of those means. "Orphaned" means are (typically) abandoned because one no longer has a larger practical reason to act on them.

belief about the constitution of such sandwiches, and his seeing the two slices of bread, he could boil that question down to whether there is peanut butter between the slices. His reasoning starts not with a *conclusion* but a *question*, and what he arrives at is a further question that, when answered, will give him the ingredients for a justification of *P*, or of not-*P*.

This kind of reasoning can be applied to technical problems, again with a change of attitude or "perspective". Agatha might reason from "How do I make it the case that there is a peanut butter sandwich?" to "How can I make it the case that there is peanut butter between these slices?", in virtue of the same capacities that Obadiah exercises, applied in a slightly different way. In doing so, she would recognize that by intentionally bringing about one circumstance she can intentionally bring about another. And while it is true in one sense that she reasons from a "conclusion", it is a voluntary one that must eventually be justified by at least one voluntary "premise".

Whether it is epistemic or technical, reasoning of this kind is hypothetical. It therefore needs to be accepted or "blessed" before it can actually justify a belief or aim. Hypothetical reasoning can prompt the immediate adoption of other new beliefs, but it would be strange to adopt a new aim in that way. There are some occasions when it makes sense to justify a new aim with an existing one, as when you add an item to your list for an already planned shopping trip. Most of the time, however, we adopt both end and means as a "package". An actual aim that *A*, rather than a hypothetical one, is a commitment to bringing about *A*. Once you establish that you could *A* in virtue of φing, you can commit to both at once in virtue of aiming to φ and taking that to technically justify your aim that *A*[12].

7.3 Relevance in Justification

In earlier formulations of the theory of technical justification I included an extra requirement that justifying aims be "relevant" to the justification. This was meant to address a potential

12. The deferred issue of partial plans will obviously require a different and more complicated answer to these questions.

for "false responsibility", where the requirement for technical warrant is met even though the agent's chosen "means" do not contribute to her end. Suppose that I know my watch is functioning well, and that it currently reads 2:58 p.m. On the basis of beliefs about the current time and my watch's good functioning, I can epistemically warrant a belief that it will read 3:00 p.m. in two minutes. But suppose I happen to aim that I tap my foot during that time. It may then appear that I can also technically warrant an *aim* that it read 3:00 p.m. by spuriously incorporating my aim to tap my foot into a technical justification.

This concern is, in an important sense, formal. Epistemic justification is usually described as a relation of beliefs to other beliefs. My model adds suspicions, and it would be straightforward to add additional attitudes such as denials or hierarchical attitudes for hypothetical reasoning (corresponding to sub-proofs in a natural deduction system). However, any such attitude would still be epistemic, and as long as one "follows the rules" and relates the attitudes in the right way, one will get a valid result. Technical rationality mixes technical and epistemic attitudes, which opens up the possibility of following those rules but still producing an invalid justification. Or, looking at the concern different way, it appeared to me that the rule that every justified aim or traim must have at least one justifying aim or traim was at best a simplified and inaccurate proxy for another much more complex rule. Not being sure what that rule was, I added an informal requirement that a justifying aim or traim be "relevant" to the justification.

I have since come to think that this addition is not necessary, and that the formal concern arises from treating syntactic proof systems as a model of justification. The "correctness" of a formal proof is a matter of syntax, and rules of the common proof systems do not enforce what I am calling "relevance". Given the right background premises, I could create a correct proof that my watch will read 3:00 p.m. that incorporates the premise that my toe is tapping, or any other premise. Anyone reading my proof would likely point out that it does not "need" or "make use of" the tapping, and that a better proof would omit that premise. However, the premise would still be part of the proof, and the proof would still be formally correct.

This is not how we generally think about justification. Suppose that Obadiah were to offer

a justification of his belief that my watch will soon read 3:00 p.m. partly in terms of the watch's good functioning, but also partly and spuriously in terms of my toe's tapping. If his argument secures the truth or high likelihood of the conclusion in the way the proof would, we *might* be happy to say that his belief is justified. However, I doubt that many would agree that the my toe-tapping is really *part* of his justification—Obadiah would just be harmlessly confused on that point. If that is right, then the concept of justification already includes a sufficient conception of relevance. It is not missing from the requirements for technical justification, but rather from formal proof taken as a model of human justification and argument.

7.4 The Phenomenology of High Technical Confidence and Voluntariness

I now turn, with some reluctance, to questions raised by the phenomenology of intention, focusing on how we might coherently experience an aim as voluntarily true.

One basis for objecting to the theory of technical justification is that it is contraindicated by, or outright incompatible with, the experience of intending. The thought would be something like this: The reasons one offers for a belief, which are presumably also what make up one's epistemic justification for it, are not other beliefs but *facts*. At the high level of credence of belief, to reason epistemically is to comprehend certain relations between facts. Because what one intends to do is, in contrast, voluntary, reasoning that relates an end to its means cannot or should not be of the same kind that relates fact to fact.

Tellingly, at least part of this objection should also apply to Cognitivism, and the way in which Cognitivist theories seem to elude the other part reveals an unresolved phenomenological tension in that approach. At first glance, the claim that an intention *is* a belief appears to bite the phenomenological bullet; if patterns of justificatory reasoning are compatible with the experience of believing, why would they be incompatible with the experience of believing in a somewhat different way? Supposing that Agatha in an important sense *believes* that she will dip, spread, and place, what stops her from *justifying* her belief that she will make a sandwich in virtue of those beliefs? The strict Cognitivist answer is that the self-referential

content of an intention-belief blocks any such inference. Indeed, the conception of an intention as "self-warranting" formally blocks its being warranted in any other way. This property has the convenient side effect that if an intention happens to get revised or abandoned for some reason, the change will not result in any epistemic inconsistency.

The unresolved tension has to do with whether and how a belief could be revised in some of the ways that intentions are revised. As I discuss in Section 3.2.1, Cognitivism provides a theory of how one can come to realize that one *cannot* φ, which would plausibly require one to abandon any intention-belief to φ, in order to avoid epistemic inconsistency. What that theory does not address is how one would stop intending what one *can* do, as far as one knows. If the worry about high-credence epistemic reasoning is that it is, or at least seems like, the relating of facts to fact, the corresponding experience of believing is presumably as of a fact. So, absent contradictory evidence, how does one abandon a fact?

This problem with Cognitivism has received relatively little attention in the literature, especially when compared to the extensive debate over whether one can rationally *form* a self-warranting belief. There are a number of arguments against revising intentions, of course, which can be read as denying that there is such a problem. However, these usually appeal either to the virtue of continence, or to the worry that agents who constantly change their minds might never accomplish anything. Those concerns have little to do with the entirely ordinary process of abandoning an intention after losing one's practical reasons to pursue it. If Agatha intends to go to the store just to buy peanut butter, not only is it rational of her not to go after discovering she already has some, it could be irrational of her to do so.

The phenomenological difficulty that Cognitivism faces is therefore to explain why, when you *believe* that you will φ as a result of having this very belief, you can just stop believing that if your motivations change. Implicit in the above objection from experience is a tie between the "factness" of believing and the epistemic norms that make that experience appropriate. If your belief is genuinely "self-warranting", and you have no other belief that is inconsistent with it, how is that warrant *epistemically* compromised by a change of practical reasons? And if it is *not* compromised, but you can revise the belief anyway, how does it count as warrant

appropriate for a *belief*? The suggestion that means-intention-beliefs might justify an end-intention-belief brings out this discrepancy. Even if Cognitivist means-beliefs are in some sense voluntary, the conception of an end-intention as a belief, together with the proposal that this belief could be justified by other beliefs, can make the whole package look worryingly involuntary.

All this is to say that the phenomenological challenge posed by intention is to combine a sense or experience of accuracy—that is, a sense that a representation matches how the world is or will be—with a sense or experience of voluntariness. And these must be combined in a way that plausibly conforms to other aspects of what it is like to intend. This need to account for a sense of voluntariness does not rest on the metaphysical issue of free will. While the nature of the human will is a matter of dispute, there is almost universal agreement that it seems to us *as if* we freely choose what to do, and that our reasoning about what to do conforms to this aspect of our experience. Even if everything one does is determined by a causal history, the bulk of that history lies outside of what one represents as reasons for acting. A theory of the phenomenology of practical reasoning must therefore account for this difference between it and epistemic reasoning. Because Cognitivism appeals only to belief, or in the case of Setiya's theory, belief combined with other properties, it has not really wrestled with this problem.

I have already described the attitude of aim as the taking of its propositional object as a "voluntary truth". In the remainder of this section I suggest how we might coherently experience an aim as both true and voluntary, and how this experience can be consistent with technical justification. As usual, I start with features of belief and epistemic reasoning, which I then adapt to serve the needs of intention[13].

13. I travel this side-path reluctantly because phenomenological arguments are famously fraught and unconvincing. Also, in the process of arguing that aim and technical justification are compatible with the experience of intending, I offer what looks too much like a positive theory. When I take a step back, I suspect that the main value of this section may be to help clarify what aspects of cognition I take to be "just" a matter of psychology. It is anyway fair to read my discussion this way, given that it is in part a response to concerns that the concerned did not take to be only about how things seem.

7.4.1 Two "Modes" of Epistemic Reasoning

It is clear that the cognitive mechanisms by which we arrive at a belief are often not evident in human "epistemic experience". From your perspective, I *believe* that the cat is not in the tree because I believe that it is on the ground. But in the first-personal moment, I typically have an experience of facts, and of relations between facts. As my credence in an epistemic attitude approaches certainty, not only does my awareness of credence recede, so does any awareness of belief itself. The result is simply an experience of the world being a certain way.

The reasoning by which we typically arrive at beliefs has a parallel arrangement; it consists of principles for understanding the world as it is. When thinking this way, the cat's being on the ground just *is* a reason that the cat is not in the tree. One awkward aspect of this experience-as-of-facts is that one can have it while being mistaken. Or perhaps such cases are merely indistinguishable: *Disjunctivism* of epistemic reasoning is the view that *facts* (ways the world is) and "facts" (inaccurate ways that one takes the world to be) are categorically different, and genuine reasoning involves only the former (Hornsby 2008). Whether or not the representations are accurate, however, to experience a belief as a fact is to experience it as infallible.

To be able to question a belief is therefore to be able to experience it differently. In my view, questioning in this way is distinct from lowering one's credence. It can *result* in lower credence, but I doubt that simply taking four beliefs to be mutually inconsistent entails first adjusting one's credence in each. The experience is instead like bracketing the attitudes in an overall sense or stance of questioning. This different perspective on belief includes a corresponding difference in the experience of epistemic reasons. Rather than experiencing some facts as revealing others, you "see" relations of justification as potentially wrong, and therefore revisable.

This dual epistemic arrangement is comparable to our capacity for vision. Most of the time, as you look around the room, your experience is just of the objects you see. You *can* attend to other aspects of the capacity—how everything is seen from a perspective, how your eyes move

and refocus—but these often recede away from attention, leaving you with just an experience of the room. From a third-person perspective, this mode of experience is a bit of a cheat. The removal, or blocking from attention, of any aspects of experience not directed towards the world can seem in the moment like the elimination of representation itself. But this cheat is presumably to our advantage. Unless and until one has evidence that a representation is not accurate, the resource-intensive "layers of indirection" that allow questioning have no cognitive role to play.

The second, questioning mode of reasoning raises the difficult question of whether experiential "directness" is inherent to a given type of reasoning, or can vary independent of it. Do we question in virtue of becoming differently aware of our reasons, or of a different, second layer of reasoning that takes the first layer as its object? According to the first possibility, what would make the difference is that "directness" would in some sense recede, leaving enough "beliefness" of the beliefs and "reasonness" of the justificatory connections for both to be questioned[14]. The second possibility is explicitly metacognitive. At the extreme, we would question evidence by forming entirely distinct attitudes *about* our beliefs and justifications. The need for a distinct second layer would suggest that the directness of "object level" reasoning is inherent: To reason from fact to fact is to set aside doubt, and therefore to doubt is to reason in an entirely different way[15].

The issue does not need to be settled to move forward. The relevant point is that we experience belief and epistemic justification in both direct and mediated "modes". The argument from the experience of belief as fact does not count against a mode of technical experience analogous to the indirect mode of epistemic experience. Of course, if the proposal were that this is the *only* way we reason technically, it would face different objections on the second, "distinct layer" theory, including the argument against routine metacognition of Section 3.4.

14. It is possible that this difference of directness is also what distinguishes lower credence reasoning. That is, we might question a belief by treating it like a suspicion of particularly high credence.

15. I am not sure that this second proposal is coherent. Even if there is a distinct second layer, its "subject matter"—that is, the beliefs and justificatory relations of the first layer—must still be open to question. I suppose it is possible that a shift to indirect thinking just replaces the first layer. However, this would make indirect thought fictional, because at the time it occurs you would lack all of the attitudes it is supposed to be *about*.

It is anyway easier to describe the experience of questioning an aim in terms of the first, "receding" theory. In contrast with a belief, an aim would be questionable in two distinct ways. The way analogous to epistemic reasoning would allow the questioning of technical confidence—of the chance that you *will* φ. Questioning confidence in a technically justified aim could include questioning the aims, traims, beliefs, or suspicions that justify it. And it could include questioning the justificatory relations between any of those. Given that technical and epistemic justification share patterns of reasoning, the basis and experience of that questioning could be the same or very similar.

The other way of questioning would relate to one's motivations for the aim, ethical or otherwise—of whether you *should* φ. In this mode, the sense or experience of voluntariness might be nothing more than an aim's being subject to this additional variety of questioning.

7.4.2 Truth, Voluntariness, and Questioning

An indirect mode of technical reasoning might be sufficient for something close to human agency. But I believe that we also often experience our high confidence intentions in a fact-like way. My claim that aim can be thought of as an attitude of voluntary truth is meant to suggest this different, direct mode of experience. This may sound like a contradiction, because *voluntary* and *fact* seem to count against one another—to conceive of a circumstance or event as both a fact and as possibly not occurring is incoherent. However, this just means that if there is an experience of voluntariness in the direct mode, it cannot be of the aim's realization being open to question.

For the sake of argument, suppose there there is *no* experience of voluntariness in the direct mode. That is, suppose that we can experience our aims as if they were facts, and our technical reasons just as if they were epistemic reasons. From Agatha's perspective, because there will be a dipping, a spreading, and a placing, there will be a peanut butter sandwich. Thinking this way, Agatha could presumably reason to a further "fact" from these and other beliefs, which she would first experience as a belief, aim, or predictable consequence upon the first subsequent switch to the indirect mode. Would this form of reasoning be incoherent?

I think that it can be coherent, as long as the shift from the direct to the indirect mode is sensitive to whether she is also conscious of reasons that call for questioning. If you are currently committed to φ in virtue of an aim to φ, what would be incoherent is to experience your eventual φing as a fact at the same time as considering your motivations for φing, or reasons that (you take to) weigh on those motivations. That is, entertaining a reason to question an aim must shift the experience of the aim to the indirect mode, so that it is "questionable". This is just as true of belief, of course. Experiencing a belief as a fact, while also entertaining a reason to doubt it, would pave the way to irrationality. An aim just has an additional set of supporting reasons, and must also be questionable when considering those reasons.

My argument that we could coherently experience aims and beliefs in the same way is not meant to suggest that we do. It is more plausible that an aim experienced in the direct mode is also marked as voluntary, but in a way not *presently* open to question. To prospectively take some φing as voluntary while not questioning whether that φing will occur is to take yourself as *responsible* for that φing. Or, if these are not entirely equivalent, a sense of responsibility is close enough to do the job[16]. The idea is therefore that Agatha would experience the dipping, spreading, and placing simply as future events in the world that she will be responsible for. And she would also experience those events, together with certain background beliefs, as reasons that she will be responsible for there being a peanut butter sandwich. In the direct mode, patterns of reasoning that "preserve" truth or high likelihood would also "preserve" responsibility.

7.5 Probabilistic Technical Justification of Traims

The "pattern" of the epistemic justification for a belief is appropriate when it secures the truth or high likelihood of that belief in relation to other beliefs and suspicions. It is in virtue of meeting this standard for high credence that the pattern can also secure the high technical

16. Interestingly, this relationship between voluntariness and responsibility is neutral with respect to tense. A past φing is unquestionable because the past is fixed, while a future φing may only be unquestioned in the moment. But one can take oneself as responsible for either in much the same way.

confidence of an aim. As I discuss in Section 5.4, there are other patterns that do not meet this high standard but still support *some* credence, and can therefore epistemically warrant a suspicion in terms of other beliefs and suspicions. The technical equivalent of *suspicion* is *traim*, and I discuss technical entitlement to traims in Section 6.4.3. It should be no surprise, then, that the last element in the model of appropriate technical confidence is the technical justification of traims, in virtue of the same patterns of reasoning that can epistemically justify a suspicion. In this section I continue to avoid much detail about the nature of lower-credence reasoning. After extending the idea of a pattern of reasoning to lower-credence attitudes, I instead discuss some examples that suggest the more general theory.

A core feature of the respective comparison of aim and traim to belief and suspicion is that technical confidence and credence are corresponding "levels" or "degrees" of confidence, which may be vague or specific. The last section's specification of a truth-preserving pattern of reasoning as fixing *content* but varying *attitude* is therefore a simplification. A belief can be justified in part by a suspicion, which has a third element: the degree of credence. Patterns of reasoning should therefore be seen as fixing not just *content* but also *degree*, considered abstractly. A 1:1 odds suspicion that a coin-flip will come up heads might therefore be replaced by a traim to get heads on a coin-flip with 50% technical confidence. The general rule is that a belief in a pattern of reasoning can be replaced by an aim with the same content, while a suspicion can be replaced by a traim of the same content and degree, as long as every aim or traim is justified in part by another aim or traim or is in virtue of a technical entitlement.

If Agatha is unsure that she will be able to spread the peanut butter—perhaps because she has never done so before, or her movement is restricted by an injury—she should lower her technical confidence that there will be a peanut butter sandwich, and therefore only traim to make one. This is comparable to Obadiah anticipating a sandwich after watching a robot attempt to make several hundred and noting that it frequently botches the spreading step. Similarly, if Agatha only suspects that there is peanut butter in the jar—perhaps they frequently run out, or are in an unfamiliar house—she should only traim to make a make the sandwich. Even though she can confidently carry out every basic action she needs to, her epistemic sit-

uation limits appropriate confidence in her end.

Aims can of course be justified by suspicions, given that they can be justified by beliefs. My belief in Section 5.4 that I will have photographed a raccoon after twenty weeks could warrant an aim to post one on my blog by then. An aim can also be partly justified by a traim. I can properly aim to get heads on a coin flip by committing to flip a coin until I do. It is not always easy to construct a natural-sounding "Obadiah analogue" for such cases. Still, technical reasoning can often be verified by replacing basic aims or traims with corresponding beliefs or suspicions, stipulating that these are warranted, and then considering what one could rightly believe or suspect on that basis. If Obadiah knows that a robot will flip a fair coin until one turns up heads, or even that there will be a series of otherwise unexplained fair coin flips until one comes up heads, his belief that a coin flip will come up heads is entirely appropriate.

7.5.1 Backup Plans

There is one special method of technical justification that is worth highlighting, both because it is common, and because it illustrates the division between technical and other practical reasons. The high confidence plans of both Carolyn and Alex in Section 2.1 depend in part on backup options[17]. To simplify slightly, a backup option is a means that *could* provide high confidence in an end, but that is less desirable than another (typically less reliable) means, where *desirability* has to do with personal value or ethics.

To simplify his case, suppose that Alex plans to either drive his own car to an AA meeting or to take a cab. His car sometimes doesn't start, but he is confident he can get a cab if he needs to. As a result, he is also confident that he will go to the meeting. At first glance, when you compare his justification for his end to an epistemic case, his high confidence may appear to be over-determined, or at least sufficiently determined by the option of taking a cab. If he is confident he can take one, shouldn't he aim to do so rather than traim, with high rather than probabilistic confidence? However, to look at his justification in this way is to confuse a

17. Alex wound up *replanning* to avoid resorting to one of his backup options, a process that is outside the scope of the this chapter's discussion of justification.

potential plan with an actual one. The probability Alex should assign to actually taking a cab is the balance of the probability he assigns to driving his own car. He should traim rather than aim to take a cab precisely because he believes there is a chance his car might start, and if it does he will drive it.

From the perspective of technical rationality, what makes a reliable backup option particularly convenient to planning is that it can fill in the balance of a probability. However vague or specific a sense you have about the potential success of plan A, your chance of pursuing plan B is the chance that A fails. If B is reliable, then your traim to A combined with your traim to B can support your end with high confidence.

7.6 "Intention"

I have already implied that the attitudes of aim and traim, together with their respective normative requirements for technical warrant, constitute a preliminary theory of intention that accounts for accurate judgments of confidence, including high confidence. The largest missing piece of the theory is the deferred account of confidence in partial plans, but such issues as acting by habit, mental actions, and intentional inaction also need to be addressed. Assuming those pieces can be filled in, however, my claim that this can be a theory of *intention* faces the easy objection that I have explicitly side-stepped the question of what the words "intention" and "intends" mean. Now that the core of the theory is in place, I will address this objection … while continuing to side-step the definition question.

As I note in Section 2.4, my worry about the focus in philosophy on concepts is the time spent debating whether some distinction X characterizes a concept C when it is clear that X is important regardless. If we *need* a theory of X (e.g. technical confidence) anyway, why not just get on with it? Our understanding of X can still be guided by C (e.g. *intention*) in a general sense, and the resulting theory might even illuminate some aspect of C. If it is unclear whether X is a joint where C cuts, and the theory of X does not resolve the question, we are no worse off than we were before—whatever issues were open just remain so.

As far as I can tell, there is only one aspect of the theory of the last two chapters that could be taken as a substantial *restriction* on the concepts *intention* or *intends*. As I note in Section 4.2, the attitude of traim necessarily associates some degree of technical confidence, however vague or specific, with its propositional content. In theory, this confidence could be as high as that of an aim, with the difference in attitude reflecting whether one was "thinking probabilistically". Given this wide range of potential degrees and specificities, what this necessity adds is something like a conceptual link between technical representation and confidence. Anscombe might make the point this way: Suppose that someone says they are going to try to φ. You ask her "How likely do you think it is that you will φ?" and she responds "How is that relevant?". If intending is inherently intertwined with an assessment of technical confidence, that answer should be difficult to interpret[18]. I see this is a feature of the theory rather than a bug, but either way it is a commitment.

The remaining aspects of the relationship of the theory to the concept *intention* can be explored in terms of cases, starting with those that are least controversial. Agatha's basic aim to spread, given its technical entitlement and eventual realization, corresponds to a straightforward case of intention to perform a basic action. Her high confidence is appropriate given her manual capacities and her warranted beliefs about her situation. That she does subsequently spread is just a further demonstration of the appropriateness of that assessment. Similarly, her aim that there will be a peanut butter sandwich corresponds to an uncontroversial non-basic intention to make a peanut butter sandwich. Setting aside whether the theory is an accurate reduction of her intention, the scenario is one in which it would be quite normal to assert that Agatha intends to make one. If Agatha had instead changed her mind about the sandwich right before acting, both of these aims would count as what Davidson calls "pure" intentions.

Sticking with appropriate high confidence, and therefore aims, what about cases in which the agent's confidence is appropriate but she is unlikely to, or definitely cannot, carry out her plan? The peanut butter jar is empty, for example, or I plan to buy a doughnut at the store

18. I take it that if she had instead merely *desired* "that φ", her answer would be understandable, if unusual: "I was just daydreaming about an idle wish. I haven't thought about it."

like I do every morning, unaware that it has slid into the sea. Such cases may raise questions about the truth conditions of "I am going to the store", but they pose no obvious problem for whether I *intend* to.

With respect to *inappropriate* confidence, however, there is more room for disagreement. At the furthest extreme, suppose that someone claims they will be on the moon Europa in two minutes—a goal that violates the laws of physics[19]. This "agent" is entirely confident, even though she cannot explain how she plans to get there. My own intuitions about whether it is really her *intention* to travel to Europa in two minutes are conflicted, and the action theory literature has some ingredients for arguing either way.

Between appropriate confidence and this extreme are a number of intermediate distinctions: The goal might be possible, and even easy, to achieve, with problem being limited to the agent's reasoning (she *could* have technical warrant but *doesn't*). Or the agent could have a plan that is completely incoherent. Or her plan is coherent, but based entirely on false information in a way that we judge makes it inappropriate—perhaps the agent is "unreasonably credulous". Or the plan would be appropriate if not for a conflicting belief the agent has failed to take into account, and that we judge she *should* take into account. Someone for whom the Europa case is not one of *intention* might put any of these conditions, or a complex synthesis, on either side of that concept's boundary[20].

Moving on to lower confidence, there is the argument, discussed in Section 2.3.3, over whether the concept *intention* includes such cases[21]. This amounts to the question of whether traims are intentions. I do not see the theory as adding anything of substance to this debate beyond providing a clear basis for the distinction[22].

---

19. Europa being more than two light minutes away from Earth.

20. Each of these distinctions is easily made in terms of the theory, and I am drawing on the theory to make them. If there are other relevant distinctions of this general kind that the theory does not "capture", those would of course count against it.

21. For brevity's sake I am assuming that these questions are separable, so that if lower-confidence technical attitudes are intentions, the various other sources of inappropriateness would change the application of the concept in the same way.

22. This issue provides an opportunity to illustrate the kind of question I am actively side-stepping. Someone

Finally, there is the question of whether intentions are always propositional. The theory already includes "broadly" technical, non-propositional representations as the analog of broadly epistemic perceptual representation. I consider it an open question whether one might pursue a basic goal in virtue of broadly technical representation without ever forming an aim or traim. If one can do so, then such representations might also be characterized as intentions.

7.7 The Three Questions

The theory of this chapter and the last adapts the theory of epistemic warrant to account for appropriate technical confidence. Rather than providing a simple summary here, I will instead briefly evaluate the theory in virtue of the three questions of Chapter 3, with the usual caveat about partial plans.

The first question is "How is technical confidence 'encoded' in, or otherwise associated with, an intention?" In the case of aim, high confidence is intrinsic to the attitude itself. The attitude of traim associates a degree of confidence, however vague or specific, with a proposi-

---

who does think that *intention* applies in lower-confidence cases, and does so in part because we sometimes say that we "intend to try", might make the following argument: "Our use of the concept *intention* in these cases implies a commonality of attitude between high and lower confidence cases. Because your theory posits different attitudes for different levels of confidence, it goes against the linguistic evidence."

This is not so much an objection to the theory of how knowledge and action are related as it is to one choice I made in illustrating that relation: My separation of aim and traim follows my separation of belief and suspicion. The latter reflects an informed preference, based on my understanding of the relevant epistemic issues. If lower confidence epistemic attitudes are instead beliefs with confidence-lowering content like "might", such beliefs would still need some variety of probabilistic warrant to be appropriate. That epistemic theory could then serve as a model for a corresponding technical theory in which lower-confidence technical attitudes are aims with confidence-lowering content.

I have found that debates in action theory based on linguistic or other evidence generally correspond to debates in epistemology based on similar evidence. These parallels count in favor of the overall relationship between knowledge and action—and more particularly the metaphorical inverse relationship—not against it. In cases where the issue is open in both realms, a substantial argument from language will have to take the form "The use of knowledge concepts suggest alternative *a*, while use of action concepts suggest alternative *b*." However, even a successful argument of this sort would only drive a wedge of that size in the parallel between knowledge and intentional action. Where the technical theory needs to diverge from the epistemic, it can.

I do not consider any theory to be "above" questions of usage. The clearest form such criticism takes are cases in which we make some distinction that a theory cannot, or cannot easily. When a theory draws a clear line that corresponds to *no* aspects of usage (including pragmatic aspects), that can also count against it. It is when we clearly do make a distinction, and the argument is that a particular concept does not, and that this is strong evidence for one underlying property or relation over others, that I am unpersuaded.

tional content[23]. Technical confidence is therefore not in virtue of the credence in some belief, but analogous to credence itself.

The second question is "What makes the encoded degree of confidence appropriate at the time the intention is formed?". The short answer is that *technical warrant* for an aim or traim makes its associated degree of confidence appropriate. In the case of a basic aim, one is entitled some movement of a body part in virtue of complex relations between "body image" and "body schema" that represent the existence and "ownership" of that part. One is entitled to a particular basic action in virtue of a capacity to assess confidence in that movement with respect to a physical scenario, such as the location of goal-related or potentially obstructing objects in relation to the body. The latter capacity can also support assessments of lower confidence, which can warrant a traim to perform the basic action.

A non-basic aim is, in contrast, technically justified by at least one other aim or traim, together with any number of beliefs or suspicions, related by a pattern of reasoning that could also support an epistemic justification. Such truth- or high-likelihood-preserving reasoning supports confidence by justifying that–given one's warranted beliefs about the likely circumstances— one's means will make it highly likely that one's end will come about. Because one's means—which are themselves either basic or justified by further means—are voluntary, so is one's end. Non-basic traims can be justified by lower-confidence-supporting probabilistic reasoning of the sort that can make a suspicion appropriate.

In addition to technical warrant, part of what makes confidence in an aim or traim appropriate is a constitutive property of those attitudes to be non-voluntarily recalled at times appropriate to act.

The third question is "How is the degree, or the intention itself, revised in light of changing circumstances, including a failure to bring about an intended means?". Revision of aims or traims for technical (rather than value or ethical) reasons is in virtue of a loss of technical warrant, which in the case of a traim could include a change in its warrant so that it only sup-

23. I take it that whether this means the confidence is part of the intention, or just associated with it, is a matter of interpretation.

ports a relatively lower degree of confidence. Such changes are analogous to what prompts the reevaluation of a belief or suspicion in light of new information. Revisions can consist of dropping the now unwarranted aims and traims, changing one's degree of confidence (including changing what was an aim to be a traim), or re-justifying the aim or traim in virtue of different means.

CHAPTER 8

Technical Warrant and Norms of Intention

The motivation for the theory of Chapters 6 and 7 is to account for appropriate technical confidence in an intention on the model of epistemology. And although the nature of action imposes distinctive features on the theory—most notably the requirement for non-voluntary recall of a basic aim—the result is not far from its inspiration in form. In this chapter I measure the theory against other norms of intention that are either not directly related to confidence, or that are not generally characterized in terms of confidence. Most of these norms are primarily associated with the work of Michael Bratman, who has explored practical reasoning in terms of requirements on intention. I also discuss some additional empirical evidence. Given that I do not alter the theory to address these norms or evidence, the extent to which the theory accounts for them provides further support for their being a close relationship between the structure of knowledge and that of intentional action.

8.1 Differentiation from Belief and Desire

One problem with the original formulation of the belief/desire pair theory (Section 3.1) is that we do not intend everything we desire. Cognitivist theories (Section 3.2) face a similar problem in that they must differentiate between circumstances we intend and those that we merely predict. A theory of intention must therefore provide some basis for distinguishing intentions from both desires and beliefs. This is a low bar, which can be met either in virtue of positing a distinct attitude or (arguably) in virtue of content added to an existing one[1].

1. Paul (2010, §3) explores one problem with theories that cannot distinguish between end-intentions and foreseeable side-effects of one's means.

In the theory of technical warrant, the attitude of aim is distinguished from desire both by being a revisable commitment to realizing the content of the aim, and by the high degree of technical confidence it associates with that likelihood. It is distinguished from belief in that the truth of the content of an aim is taken as voluntary rather than involuntary. Traim is distinct from suspicion for the corresponding reasons, and from desire in virtue of its necessary association with its content of some degree of technical confidence, however vague or specific[2].

8.2 Means-end Coherence

Bratman (1999, p. 31) includes this passage:

> Second, there is a demand for *means-end coherence*. Although plans are typically partial, they must still be appropriately filled in as time goes by. My plans need to be filled in with subplans concerning means, preliminary steps, and relatively specific course[s] of action, subplans at least as extensive as I believe are now required to do what I plan. My plan to go to Tanner Library will need at some point to be filled in with a specification of a means to getting there (car? bus? bike?) and may also need to include a specification of appropriate preliminary steps (for example, checking the bus schedule). And my plan to go to a concert tonight needs at some point to be filled in with a specification of which concert. Failure to fill in my plans as needed in these ways will leave them means-end *in*coherent.
>
> Of course, means-end coherence does not require that my plans specify what I am to do down to every last physical detail. Rather, my plans will typically be at a level of abstraction appropriate to my habits and skills. My plan to take the bus to Tanner Library need not include a detailed specification of the foot with which I am to step onto the bus when it comes. ... means-end coherence does require that my plans be filled in with specifications that are as detailed as is, on my view, needed for their successful execution. That is not to say that they need to be filled

2. The question of what might constitute a "distinct attitude" is an interesting one. Suppose (against evidence) that the only remaining objection to an otherwise Cognitive approach to intention was the lack of a distinction between intention and belief. Realizing this, someone proposes that intentions are "ileafs", which are subject to all the same requirements as beliefs and can serve in any context a belief can. The belief/ileaf distinction would be tracked by the agent, but would have no other consequence. I am not sure whether this proposal is coherent. I see attitudes as sets of characteristics that we associate with propositional contents, and the terms "belief", "desire", and so on are ways we talk about those sets. Regardless of whether a mere tag is sufficient, aim and traim qualify as distinct attitudes by the stronger standard.

> in all at once; it is enough that they be sufficiently filled in before it is, by my lights, too late[3].

Some of this "demand" has to do with partial plans, and our capacity to plan over time —a subject I am deferring and therefore will only partly address here. Even so, it can also be read as a requirement on the totality of one's planning, evaluated after either one acts intentionally, or attempts and fails in a way that could have succeeded. That totality—setting aside the possibility of replanning as one goes along—is within the scope of the requirement for technical warrant as it now stands. In warranting an aim, one meets a stronger requirement than means-end coherence; the former implies the latter.

More specifically, technical justification gives form to the means-end relationship in relation to one's warranted beliefs. To rationally take *M* as a means to *E* is to understand that *M*,

3. This early formulation of the requirement has since been overshadowed by one of questionable relevance. In Bratman (2009, p. 13), he describes a different norm that he calls the "instrumental rationality requirement" or IR: "Suppose I intend end E, believe that a necessary means to E is M, and believe that M requires that I intend M. My attitudes concerning E and M engage a basic requirement of practical rationality, a requirement that, barring a change in my cited beliefs, I either intend M or give up intending E." About this new requirement he says, "In other work I have focused on what I have called a requirement of means-end coherence of one's intentions and plans. I see IR as a central aspect of that requirement, though the requirement of means-end coherence goes beyond IR, strictly speaking ... But IR is at least the heart of the requirement of means-end coherence, and it will simplify my discussion here to focus on it."

The now-common idea that this new requirement is the "core" of means-end coherence is dubious. For one thing, although Kant talks of necessary means, the idea that we frequently encounter and recognize a necessary means in daily life has little support; As Vogler (2009, p. 154) notes, there are usually many available ways of pursuing an end. IR is also unusually meta-cognitive in comparison with Bratman's other views. It is important to plan in time to act, but the idea that one routinely does so in virtue of a belief *that φing requires that [one] intend to φ* strains intuition. Regardless, other authors, including John Brunero, have gone further than Bratman and simply relabeled IR as "Means-Ends Coherence" (Brunero 2010, p. 580). Worse, other authors now drop the additional belief, which at least has the benefit of relating the requirement to planning before it is too late[†]. According to M. Schroeder (2009) (and assuming partial plans are otherwise acceptable), if I happen to plan to intend that E and notice (for whatever reason) that M is a necessary means to E, I must immediately make that aspect of my plan *un*partial. Why non-necessary means can wait while necessary means cannot is left unexplored[‡]. All this is to say that the general discussion of means-end coherence has taken what I consider to be an unfortunate detour.

For what it is worth, the no-partial-plan specification of technical warrant clearly does require that one justify an aim or traim *E* with what one takes to be a necessary means *NM*: to fully justify *E* without *NM* is to conceive of another justifying aim as a means to *E*. The remaining question is whether necessary means have a special relationship to "partial planning". I doubt that they do.

[†]I am interpreting the role of the belief *that φing requires that I intend to φ* as requiring me to intend to φ *now*. In essence, I say to myself "Oh dear, I'm running out of time to intend my necessary means, so I had better start doing that." Otherwise, that belief would be a strange epicycle on top of already believing that my means is necessary to my intended end. Bratman himself cites Binkley (1965, p. 443) as explaining the need for the extra belief; I read Binkley as being concerned about a different issue.

[‡]Raz (2005a, pp. 5-6) makes this argument in more detail.

together with one's beliefs about the likely circumstances, secures the truth or high likelihood *E*. This "test" can be applied just as well in the context of a partial plan. Suppose you have not yet decided how to pursue *M*. You should take *M* as a means to *E* only if *M* can technically justify *E*, and meet whatever further requirements govern as-yet unplanned aims.

The lower-limit on planning that Bratman describes in his second paragraph also matches the reasoning criterion for basic action. This is not surprising, and not evidence of much. Advocates of the movement criterion recognize that we do not consciously work out every individual movement. Such unconscious reasoning could therefore not be part of planning as Bratman conceives of it. The theory of technical entitlement, however, does add something. For "habits and skills" to be the foundation of confident plans, we must have a basis for confidence in intentions on that level. What is a skill, or some *degree* of skill, can be accounted for in terms of entitlement to an appropriate degree of technical confidence given the likely circumstances of acting (or attempting). For a basic aim or traim *B* to be a means to *E*, it must technically justify it. The appropriateness of that relationship will depend in part on whether one's degree of confidence in *B* corresponds to one's entitlement.

8.3 Consistency Requirements

As I discuss in Section 2.3.3, Bratman limits the application of *intention* to plans that are not subject to certain consistency requirements. I also argue that his distinction between intentions and other plans corresponds to the distinction between high and lower confidence. If both of us are right, then aims are subject to the other requirements described in Bratman (1999, p. 31):

> To coordinate my activities over time a plan should be, other things equal, *internally consistent*. Roughly, it should be possible for my entire plan to be successfully executed. Further, a good coordinating plan is a plan for the world I find myself in. So, assuming my beliefs are consistent, such a plan should be consistent with my beliefs, other things equal. Roughly, it should be possible for my entire plan to be successfully executed given that my beliefs are true. This is a demand that my plans be *strongly consistent, relative to my beliefs*. Violation of either of these consistency constraints tends to undermine the contribution of my

> plan to coordination in the world I am in. If I plan both to leave my car at home for Susan and also to drive my car to Tanner Library, all the while knowing that I only have one car, I am unlikely to succeed in my effort at coordination.

This sort of consistency corresponds to the first pre-condition on the shopping in Anscombe (1957, §32): "For if the discrepancy [between the list and what the man buys] might arise because some of the things were not to be had and if one might have known they were not to be had, we might speak of a mistake (an error of judgment) in constructing the list."

The theory of Chapter 7 only imposes requirements on aims in means-end relations, rather than on pairs of otherwise unrelated aims, or aims and beliefs that have no justifying relationship. Even so, some coherence between aim and belief may already be addressed by epistemic requirements. Suppose that Obadiah believes there is almost always peanut butter in the house, but also has evidence that there is no peanut butter today. In that case he should not justify a belief that there is peanut butter now in virtue of his general belief alone, while ignoring that evidence. A pattern that is inconsistent in an epistemic context should also be so in a technical one, and Agatha should not aim to make a peanut butter sandwich under the same circumstances. Still, this caveat does not obviously extend to reasoning among aims. If Bratman is right, some further requirement is needed.

Happily, that requirement just the analogue of epistemic consistency (Section 5.6), and therefore not an addition to the theory but just a subject I have put off until now. Belief is subject to certain consistency norms. As the analogue of belief, we might also require that aims be consistent with one another. As norms go, however, this is a little vague. Inconsistency goes beyond direct contradiction, such as aiming to go to the store and aiming not to go to the store. It also applies to aiming to go to store at three o'clock and simultaneously aiming to go to the library at three o'clock. And there is still the question of "strong consistency" in the general case.

As it turns out, it is easier to address Bratman's consistency requirements in terms of a relation of *capacities* than a relation of *norms*. In Section 1.2.2 I discuss how a capacity to "detect" inconsistent beliefs might be accounted for in terms of a capacity that approximates epistemic closure—presumably biased towards those further beliefs more likely to be useful

in daily life—and a capacity to detect direct contradictions. This is not a proposal, just an illustration of how two capacities in combination might account for another. Along similar lines, if humans are subject to norms of epistemic consistency (within certain limits), we presumably some capacity to detect inconsistency so that we can take steps to eliminate it. Given that we are also required to warrant beliefs, this would be a capacity to detect inconsistency among beliefs *related to one another in networks of justifying relations*. Even if Q does not partly justify R, nor R Q, Q might partly "justify" P and R might partly "justify" not-P for someone temporarily suffering from inconsistency. Such relations, or others that trace backwards or forwards though the "glue" of justification, are the most likely basis of our capacity to detect inconsistency among beliefs.

Belief, however, is just an attitude towards a proposition. Assuming we have this capacity for resolving epistemic inconsistency, we can also think of it as a capacity that detects *propositions* with inconsistent contents, when those propositions are arranged in justifying relations. The role of the attitude of belief in relation to this capacity would be to pick out those attitudes that subject to the consistency requirement. So suppose that instead of "operating on" beliefs alone, the capacity also targeted aims. Aims, after all, are also arranged in justifying relations of the same patterns, including those that inter-relate them with beliefs. The operation of the capacity would then enforce this rule: If all one's present aims were swapped with beliefs of the same content, and a swapped belief is inconsistent with another (swapped or not), then the original aim is inconsistent in the analogous way[4].

Although this description is thick with metaphor, the point it makes can stand on its own. Given the requirement for technical warrant, non-basic aims are interrelated with beliefs and with each other much like beliefs are interrelated. If whatever capacity accounts for our ability to avoid incompatible beliefs (to the extent we can) were extended to aims, it would meet both of Bratman's requirements (to the extent we can). If the theory is wrong, and our means-end reasoning is not in virtue of the same patterns of reasoning that apply in epistemic justification,

4. Basic aims may need some special treatment in the analysis, given that the contents of some may be incompatible with epistemic norms, and therefore cannot just be "swapped".

strong consistency would need to be explained in some other way.

8.3.1 Aiming While Believing

There is another tricky form of consistency not emphasized by Bratman, which is that one should not typically aim that P while also independently believing that P. While it is also unusual to traim that P when one believes that P, doing so can be perfectly reasonable. You can traim to win a contest, for example, while confident that someone will win[5]. Aiming that P while believing that P is typically irrational, however, unless you take the latter to be in virtue of the former. The underlying requirement, which is difficult to state, comes from the intersection of overall consistency with the relevance that I discuss in Section 7.3. A technical justification is not just improper if one could remove the justifying aims while leaving the conclusion intact. It is also improper if one could epistemically justify the content of the aim with other of one's beliefs. This is equivalent to an epistemic mistake of believing for the wrong reasons. Q and R might normally be reasons to believe that P, while not being so right now, when S and T account for it. Taking P to result from your basic actions, rather than other factors, is an analogous mistake.

8.3.2 Probabilistic Consistency

Bratman's consistency requirements are also his basis for rejecting what he calls the "Simple View"—that everything we do intentionally is preceded by an intention to do it. Setting aside the conceptual claim, his famous video game argument is in virtue of a plan that violates strong consistency (Bratman 1984, pp. 381–3):

> ... let us turn to a series of three examples. In the first case I am playing a video game in which I am to guide a "missile" into a certain target. I am quite skilled at such things, but it is a difficult game and I am doubtful of success. Still, I aim at the target and try to hit it. As it happens, I succeed in just the way I was trying. ...

5. Of course, what you really want is that *you* win the contest. However, it is not necessary to conceive of your goal that way, unless there is some risk of accidentally helping some other player, which you hope to avoid.

> Do I hit the target intentionally? It seems that I do. …
>
> Suppose now that a second game is added, a game which also involves guiding a "missile" into a certain target. Since I am ambidextrous and can play one game with each hand, I decide to play both games simultaneously. As before, the games are difficult and I am doubtful of success at either of them. As it happens, I miss target 2 but do succeed in hitting target 1 in the way I was trying and in a way that depended on my relevant skills. Here again, it seems to me, I hit the target intentionally.
>
> Let us now suppose that the two games are known to me to be so linked that it is impossible to hit both targets. If both targets are about to be hit simultaneously the machines just shut down. Both targets remain visible to me, so I can see which target I hit if I hit either target. And there is a reward for hitting either target. But I know that while I can hit each target, I cannot hit both targets. Still, I know it is difficult to hit either target, so I again decide to play both games simultaneously; I see the risk of shutting down the machines as outweighed by the increase in my chances of hitting the target. I proceed to try to hit target 1 and also try to hit target 2. I give each game a try.
>
> Suppose I do hit target 1 in just the way I was trying to hit it, and in a way which depends heavily on my considerable skills at such games. It seems, then, that I hit target 1 *intentionally*. So, on the Simple View, I must intend to hit target 1. Given the symmetry of the case I must also intend to hit target 2. But given my knowledge that I cannot hit both targets, these two intentions fail to be strongly consistent. Having them would involve me in a criticizable form of irrationality. But I seems clear I need be guilty of no such irrationality: the strategy of giving each game a try seems perfectly reasonable.

Bratman describes himself during these games as "doubtful of success" and as "trying to hit" the targets, both marks of what I call lower technical confidence. According to the theory of technical warrant, therefore, Bratman traims to hit each target with the degree of technical confidence to which he is entitled by his skill. If the consistency requirements of aim correspond to those of belief, then the consistency requirements of traim can correspond to those of suspicion. This further adaption of the epistemic model conforms to Bratman's claim: Given sufficient familiarity with Bratman's level of skill, Obadiah could rightly suspect that Bratman will hit target 1, and also suspect that he will hit target 2, while appropriately believing that he cannot hit both targets.

However, this negative claim about propositional consistency is rather trivial compared to what else Obadiah could believe in virtue of some probability theory. Suppose that Obadiah

also appropriately believes that Bratman is truly ambidextrous. He can then reason that the upper limit on the probability of Bratman's scoring is 0.5: If Bratman generally hits the targets with 50% accuracy, the probabilities that he will miss both targets, hit target 1, hit target 2, and that the game will shut down are each 0.25. If Bratman instead appropriately tends to hit targets with either greater or lesser accuracy, the chance he will hit only one will go down from there.

Across all of human cognition, probabilistic calculations of this sort are probably uncommon; it is more the kind of thing one learns at school than a product of intuition[6]. Even so, there does seem to be a general human capacity for simpler forms of *probabilistic consistency*. To give a trivial example, Obadiah should not assign any more or less epistemic confidence to Bratman's scoring in the single missile game than he does to Bratman's hitting the target. And he should not assign lower credence to the prospect of at least one of two coin flips coming up heads as he does that the first will. Given the relationship of traim to suspicion, this consistency can easily be adapted to the technical realm. It would prohibit Bratman from traiming to score in the first game with different confidence than he traims to hit the target. And it would prohibit me from traiming to get heads at least once in two coin flips with less confidence than I traim to on the first one.

I think it is clearly true that lower confidence intentions are subject to this loose sort of consistency, in at least rough proportion to our sense of consistency among lower-credence epistemic attitudes. Accounting for it in virtue of probabilistic technical warrant is just a matter of a further generalization: As beliefs are subject to one kind of consistency, suspicions are subject to another, including requirements for consistency between our suspicions and our beliefs. The technical justification of a traim is in virtue of the same patterns of reasoning that can justify a suspicion. Therefore, our capacity for joint consistency of technical and epistemic attitudes can be accounted for simply by expanding the scope of the already-presumed capacity

6. This is not a general claim about human neural function. There is evidence for routine probabilistic calculations in specific cognitive sub-systems. Some of aspects of perceptual representation, for example, are in virtue of neural processing that some claim correspond to Bayesian functions. I am only speaking here of conscious reasoning, or chains of reasoning available to consciousness.

that ensures joint consistency of suspicions and beliefs[7]. Once again, there is nothing more to explain[8].

8.4 Commitment

Bratman also characterizes "future-directed intention" in terms of two dimensions of *commitment*. The first dimension is *volitional* (Bratman 1999, p. 15–6):

> Both intentions and desires are, but ordinary beliefs are not, *pro-attitudes*. Pro-attitudes in this very general sense play a motivational role: in concert with our beliefs they move us to act. ...
>
> Intentions are, whereas ordinary desires are not, conduct-*controlling* pro-attitudes. Ordinary desires, in contrast, are merely *potential influencers* of action. The volitional dimension of the commitment involved in future-directed intention derives from the fact that intentions are conduct-controllers. If my future-directed intention manages to survive until the time of action, and I see that the time has arrived and nothing interferes, it will control my action then.

As should be clear from my account of the role of the will in Section 6.6.1, with its conception of agent-centered action, this is not the theory I prefer. But it is close, and the attitudes of aim and traim provide a similar kind of commitment. While they do not *control* an agent's

7. I suspect that, to the extent that we judge that *basic* aims and traims are mutually consistent, it is in largely virtue of imagining a scenario in advance. Just you should not have high technical confidence that you can move your arm through a table, you should not be confident that you can move your right arm to the left, and your left arm to the right, if their paths would intersect.

8. I try in this thesis to note the relevant experimental evidence I have encountered, whether it counts in favor of the theory or against it. Still, this is not a work of experimental philosophy, because none of its claims or implications have been tested by further experiments. However, this idea that the same mechanism underlies consistency of belief (and suspicion) and consistency of aim (and traim) provides one opportunity for such a test.

We likely detect different kinds of inconsistency at different average speeds, and with different average accuracies. The speed and accuracy with which we judge inconsistency of an aim with another aim or belief can provide evidence for similarities or differences of underlying representation. Suppose, for example, that our reasoning about technical confidence is routinely meta-cognitive after all. The psychological content of that reasoning likely makes the detection of some kinds of inconsistency more difficult than if it were in virtue of "ground level" cognition. The underlying representation of our intentions might therefore be tested experimentally by comparing human speed and accuracy of various intention inconsistency scenarios with analogous epistemic scenarios. In one of the latter, a participant would be prompted to detect possible inconsistency among "ground-level" beliefs. In the other, the beliefs would include some psychological components of the sort supposed in Inferentialist theories. The mean and distribution of speed and accuracy across the different tests could provide evidence in favor of one of the theories.

conduct, they serve as a record of one's commitment to act. Awareness of that commitment at the time appropriate for acting allows the agent to fulfill her commitment.

The second is a *reasoning-centered* dimension, which Bratman characterizes in terms of "two main facts". The second fact has to do with the role that a partial intention plays in prompting one to plan—a topic I once again set aside. The first fact "is signaled by talk of being *settled* on a certain course of action" (Bratman 1999, p. 16–7):

> My intention resists reconsideration: it has a characteristic *stability* or *inertia*. …
> Of course, given new information, or a change in what I want, I may well reopen the question and reconsider. As Austin emphasized, "Every intention … which regards the future, is ambulatory or revocable." But revocability does not entail actual reconsideration. Lacking new considerations I will normally simply retain my intention up to the time of action. Retention of my prior intention and nonreconsideration is, so to speak, the "default option."

There are really two standards here, and much of Bratman's subsequent work in action theory rests in part on the stronger one of *resistance* to reconsideration. The minimal conception of aim as a voluntary truth (or traim as a voluntary increase in likelihood) involves a weaker commitment. Whatever reasons originally counted in favor of adopting an aim or traim should—"lacking new considerations"—count against revoking it. If there is additional weight in its favor, my preference would be to account for it in terms of the investment of cognitive resources in arriving at the decision[9], the potential costs of forming a new plan, and the risk of not planning in time to act. While that risk is a technical one, the actual and potential cognitive costs have more to do with value, and are best reasoned about in that realm. I therefore prefer his "default option" standard, which corresponds to the requirement on aim and traim that one either attempt to act or change one's mind.

8.4.1 Motivation and Recall

Related to the commitment involved in intending, there is one way in which I consider the theory of technical warrant (so far), and also Bratman's analysis, to be too optimistic. Settling

9. See also Section 9.4.3.

on a course of action is of little help unless one will eventually have the opportunity to act, given the right circumstances. With aim or traim, that opportunity is supposed to be in virtue of the commitment involved—the attitude should persist until either the time to act arrives or one changes one's mind—and non-voluntary recall. If an agent arrives at an aim that is not well supported by her further practical reasons, she might fail to act due to weakness of will (Section 6.6.1). But weakness of will is not appropriate. It implies a mistake of reasoning, and does not obviously pose a challenge to norms of appropriate technical reasoning.

What I worry is too optimistic is the view that intentions of middling motivation—which Shepherd (2017) discusses as driving "half-hearted action"—are reliably recalled. I have not yet turned up experimental studies that isolate this question, which anyway seems difficult to investigate. In my own experience, however, the likelihood of my even recalling an intention is in rough proportion to my sense of its practical importance. This makes some sense in terms of the allocation of cognitive resources. Non-voluntary recall is not "free"; to remember an intention at the right time requires picking out its associated condition. It is obvious that those intentions that are most important to us are *exciting*, making it difficult to think about anything else, or to sleep and risk missing our opportunity[10]. It would be consistent if lukewarm intentions were granted few cognitive resources, and therefore only a chance of coming back to mind at the right time.

In the end, though, forgetting a middling intention is not necessarily a failing of technical rationality per se. What is not very important to us is not very important to us. When one has barely "settled" on what to do on a practical level, one's ambivalence might leak into the technical one. This would just mean that we lack a general capacity to overcome the force of our practical reasons by stipulation.

10. I think importance, rather than desirability, is the most relevant criterion here. This phenomenon should probably be distinguished from other kinds of anticipation, which can overlap. I may not be able to sleep because I am going to be tortured tomorrow, although I have no active role in the process. If I am instead being sent into battle, my excitement will likely be due to both my role in the battle, and my anticipation of a stressful and unpleasant situation.

CHAPTER 9

Further Practical Rationality

I argue in Section 2.2 that technical judgments are separable from other aspects of practical reasoning. In planning, the likelihoods of your means bringing about an end, and of your attempt to bring about that means, can be assessed without regard to what you *should* do. And what a person has intentionally done, or has failed to do, depends only on what she intended and what has happened, and not on what she should have done. Part of what makes the term "technical" appropriate is that confidence in our intentions is in virtue of a distinct *realm* of reasoning, as assessments of credence are in virtue of the realm of epistemic reasoning. As a basic principle, this separation is at least implicit in much action theory going back at least as far as Anscombe. Key arguments of *Intention*, including its characterization of the A-B-C-D order of the pumping example, depend only on technical analysis rather than claims about what an agent desires.

Even so, the distinction does not seem to have had much influence on ethical philosophy. Perhaps this should not be surprising, given that the separation of technical reasoning from other sorts has no obvious ethical application. What one should do clearly depends both on what one values and on what one can do. Even if what is technical can be specified separately from what is ethical, the latter cannot be separated from the former. Ethical reasoning depends on some conception of instrumental reasoning, and ethicists frequently theorize about technical problems. As a result, contemporary ethical theories of the relation of means to ends are partly integrated with contemporary action theory and partly independent of it.

In this chapter I explore some implications of separable technical rationality for ethical theory. I have argued that the attitudes of aim and traim, with their respective normative

requirements, have a number of advantages as an account of intention. Those requirements are separable in that they do not depend on considerations of ethics or value. What they are separable *from* I will call "further practical rationality". This includes all cognition, including but not limited to reasoning, relating to assessments of value and morality, and the influence of those assessments on decision-making. The overall viability of aim and traim as a basis of intention, and therefore intentional action, depends in part on whether they can serve as part of a theory of further practical rationality. Whether or not they have obvious ethical *application* is not my concern; my interest is in what it means for ethical reasoning to depend on *this* conception of instrumental reasoning, rather than another one.

This is not just the question of whether aim and traim can simply be "slotted into" existing ethical theories. That would be a strange standard, given that members of the set of *ethical theory* are incompatible, and generally all over the place. And I doubt that it is the question of *which* existing theories they can be slotted into, and whether those are more promising by other standards. I am not an ethicist, but after familiarizing myself with the more common conceptions of instrumental reasoning, I believe that it is too early for unification with current ethical theories to count as evidence for or against technical warrant, because its implications have not yet been sufficiently explored.

Still, separability itself clarifies the ground of those implications. If technical warrant does not depend on further practical reasons, but further practical judgments depend in part on technical judgments, then technical rationality is *prior to* further practical rationality. The problem of reformulating ethical theories in terms of aim and traim can largely be reduced to the problem of confronting and addressing this relative priority. Most of this chapter is an attempt to describe the latter problem, and how ethical conceptions of instrumental reasoning have not fully grappled with it. I also come to two tentative conclusions. One is that separability does not weigh in favor of one broad approach to ethics (such as Consequentialism or Deontology) over another. The other is that technical priority may resolve, or perhaps dissolve, much of the debate over the nature of so-called "instrumental principles".

9.1 Reasoning from End to Means

The subject of technical rationality most often arises in contemporary ethical theories under the guise of *instrumental reasoning*. This is a *process* of reasoning that starts with an end and produces an appropriate means to that end. Stated this way—and it *is* often stated roughly this way—the concept of instrumental reasoning is already somewhat question-begging. If my goal is to eat a dodo sandwich this afternoon, and then I discover that dodos are extinct, it is questionable whether that goal ever really had the status of an *end*. Alternatively, if a dodo couple had recently been discovered I might similarly see that although I *could* eat a dodo sandwich, the chance of propagating the species unquestionably outweighs a culinary preference. In either case, reasoning that was supposed to be merely "instrumental" winds up violating the premise that what I represent is my end. At best it *was* my end, and more plausibly it was a desire that I evaluated as a hypothetical end.

In addition to finding that there are no means to some hypothetical end, or that all means are inappropriate, I could find that even the best available means only supports lower confidence. Deciding that I want a peanut butter sandwich instead, I could discover that I can only make my eating one a third more likely. I may then make that my end or not, depending on how I weigh the potential costs, risks, and benefits. As in the other cases, supposedly "instrumental" reasoning that only yields lower-confidence means can undermine the status of a goal as an end.

The grain of truth in this conception of instrumental reasoning is that we do often choose potential ends partly in light of a prior sense that we can arrive at satisfactory means to them. However, if the reliability of our prospective sense of technical confidence in such cases depends on technical rationality, it is strange to limit the scope of "instrumental reasoning" to only those cases already worked out (in some sense) in virtue of broadly technical cognition. If the term just describes those cases in which an agent's prospective goal turns out to be her actual one, then it does not describe a process so much as one potential result among others. If it instead describes those cases in which our prospective sense of technical confidence is

validated, then it likely picks out the later stages of a larger technical reasoning process. Neither of these interpretations cuts at a joint, or matches ethical conceptions of "instrumental rationality". Way (2012), for example, begins, "We are instrumentally rational when we take necessary and effective means to our ends, and instrumentally irrational when we fail to do so."

Way's article also includes a summary of various instrumental *principles*, starting with this "simple" formulation[1] (§2):

> (Simple Principle) If you intend to *E* and believe that *M*ing is necessary for *E*ing, you ought to *M*.

The now-standard objection to this proposal is that it sanctions "bootstrapping": Given that whether you ought to *E* is left open, the requirement raises the status of *M* as something you ought to do merely because you intend *E*. The implication is that if you take *E* as a prospective goal, but then discover that you can only bring it about by doing something reprehensible, you should go through with the plan anyway. If "ought" can be used this way, it is not in the sense that some action is ethically appropriate or required.

Korsgaard suggests this alternative:

> (Transmission Principle) If you ought to *E* and *M*ing is necessary for *E*ing, then you ought to *M*.

And others have suggested roughly this one:

> (Wide-Scope Principle) You ought (to *M*, if you intend to *E* and believe that *M*ing is necessary for *E*ing)

The latter being consistent with either *M*ing or dropping your intention to *E*[2].

At first glance, the theory of technical warrant might appear to favor the Wide-Scope Principle over the others. Suppose that I aim that *E* by means of *M′* but then learn that *M′* will not secure the truth of *E* after all. If the only other way of realizing *E* that I am aware of is *M*,

1. I am italicizing goal symbols in this and the other quoted material of this chapter for stylistic consistency.

2. The Simple Principle could also be called a "Narrow-Scope Principle".

then to conform to technical requirements, I must either justify E by M or or stop aiming that E. However, technical requirements have nothing to do with what you ought to do in a further practical sense, which is how the "ought" in the Wide-Scope Principle is usually read. And anyway, both the Simple and the Transmission principles conform to technical requirements just as well. Technical rationality relates only to what you will do, are doing, or have done, or hypothetical reasoning about such, and cannot favor any further principle about what you should do over any other.

9.2 Confusing Process with Product

The usual worry about the Wide-Scope Principle is that it negates the commitment of intending, in that it does not "weigh in" on whether you should continue to intend to E. This objection reveals how much is being asked of these principles. Practical rationality is not supposed to be about what one should "do", in the sense of staying consistent with some formal disjunction. It is supposed to be about what one should *do*, in the sense of what to take as one's end. The idea that an *instrumental* principle can specify what to *do*, even in the sense what to take as a means to an end, rests on the same mistakes that lead to the faulty picture of instrumental reasoning as starting with an end.

I suspect that this picture comes from confusing the temporal order of steps in a reasoning process with their priority. Consider this principle of epistemic reasoning based on a loosely analogous confusion:

> (*A Priori* Principle): A belief is *a priori* only if it is warranted in virtue of no empirical beliefs

One problem posed by this principle is how to interpret "in virtue". Suppose that you believe Pierre Wantzel's proof that one cannot trisect an arbitrary angle with a straight edge and compass. If you *read* his proof, then your belief is plausibly "in virtue" of your perceptual beliefs about what you read. And if you learned the proof for a test, your belief is partly "in virtue" of a belief that it might be on the test—an event that counts as empirical. On that

reading, however, the principle is false, because those empirical influences on reasoning do not preclude a belief's being *a priori*. What makes a belief *a priori* is a quality of its epistemic warrant—whether its justification ultimately depends on the truth of some empirical fact. The process by which you arrive at that justification is beside the point. The phrase "in virtue" is therefore too vague; either a different term or a very careful clarification is needed.

The faulty picture of instrumental reasoning likely comes from similarly confusing a typical sequence of steps taken in a reasoning process with a pattern of justification. The analogous thought would be, "If I start with an end, and use facts about my skills and situation to arrive at a means, then the end and those facts must determine my means." Something like this thought lies behind the ethical concept of *transmission* of reasons for an end to reasons for a means, as suggested by this Facilitative Principle (Raz 2005b, pp. 5–6):

> When we have an undefeated reason to take an action, we have reason to perform any one (but only one) of the possible (for us) alternative plans that facilitate its performance.

The problem with this principle is not that it is wrong, but that it is right only in virtue of an obscured circularity. An *undefeated* reason is not (only) a *weak* reason. You have a weak reason to wipe your runny nose when a child is drowning nearby, which is defeated by the priority of saving her life. An undefeated reason is one not defeated by any other, and therefore a reason for acting. The Facilitative Principle and Korsgaard's Transmissive Principle make it appear that one can *start* with an undefeated reason for an end and *move* to a reason for a means, perhaps leading to a competition among potential means from which one will emerge undefeated. Because I ought to pay my mortgage, for example, I ought to sell my boat when that is my only means of doing so. Given my predicament, my reasons for paying my mortgage are reasons for selling my boat.

Assume, for the sake of argument, that in this case my technical reasoning is sound. I could possibly raise money in other ways, such as winning the lottery or patenting a brilliant idea I could immediately sell. I could also resort to stealing or blackmail. But as it happens I can do none of these things with high confidence. If I am going to pay my mortgage on time,

I need to sell my boat. Unfortunately, it eventually becomes clear to me that the only person who may buy my boat is my credulous cash-rich friend, and that it will happen only if I lie about what it is worth. Is my undefeated reason for paying my mortgage then a reason to lie to my friend? Most likely not—not making the payment is probably the better option. This result is entirely consistent with the Facilitative Principle—it just turned out that my reason for making the payment was defeated after all.

What this example shows is that any *local*[3] theory of transmission is unworkable. If there is "transmission" it is like that of electricity—a closed circuit from an end all the way through to basic means is needed for any reasons to "flow". One might just as well "reverse the polarity" and say instead that reasons transmit from means to ends: You should *E* because you should *M*. Reversing the principle to this extent takes some imaginative effort, but a less ambitious version matches simple intuition:

> (Reverse Transmission Principle) If *M* is the only available means to *E*, then you ought to *E* because it is not the case that you ought not to *M*.

9.3 The Priority of Practical Reasons

The illusion of local transmission is a misinterpretation of one way we reason more efficiently. This improvement is vital to our ability to make practical decisions, but does not alter our rational basis for them. As a point of comparison, consider the problem of deciding on a strategy for the next three moves in a chess game. The set of legal moves is determined by the identity and placement of the pieces and the next two moves of your opponent, which you do not yet know but might predict. What makes one move better or worse than another is how it contributes to your winning or drawing the game. Chess strategy importantly depends on what moves are legal, but not the reverse. Even the winning condition lies just beyond the rules that govern the king—all they tell you is that that piece must move and there is nowhere to move it to.

3. That is, pertaining to a single or a fixed number of means-end relations.

One method of selecting moves would be to start with a list of every legal sequence of five moves—three of yours and two of your opponent's. Given a way of judging, for a given arrangement of pieces on the board, which player is in a better position after a move, you could predict the more likely opposing moves given your belief that your opponent also wants to win. The reason this approach is ruled out in practice is because it is *intractable*[4]. Before the "end-game", at least, there are too many sequences of moves to remember and consider, given the time allowed to settle on a move.

A more realistic method is to reject those first moves that already seem to leave you in a worse position, and focus on a few that seem most promising. Then you can judge your opponent's most promising responses to those moves, and go on from there. By strategizing this way, you do risk never considering the best combination, because what looks worse after one move might look better after three. But any tractable strategy poses that risk, and that limitation is part of what makes the game challenging.

What is important to see about this second method is that can never yield a better option than the first. When you are limiting yourself to considering three moves and two counter-moves, the set of five-move sequences, as determined by the rules of the game, includes every possibility. Ideal strategy is also determined by those rules, in conjunction with your opponent's ability. No method of limiting the moves to consider can alter the relative priority of the legality of moves in relation to strategy.

The illusion of local transmission comes from analogous methods for arriving at a plan. The full set of outcomes one might bring about with *some* confidence—all of the potential aims and traims one could warrant—correspond to the three moves. The two counter-moves are contingencies—circumstances you cannot predict with high confidence, which may partly depend on what one does first. A person cannot, and therefore does not, consider more than a tiny fraction of these possibilities in deciding what to do. But the fact that one typically begins by considering a few potential ends that seem most promising does not alter the relative priority of technical to further practical reasoning. If you could consider every individual technical

4. For most people, at least. A computer can definitely "look" three moves ahead in a chess game.

possibility—"bundles" of end, specific means, and likely non-intentional consequences—you could rank them on the basis of ethics and value and pick the best. The human process of decision-making is an extreme approximation of that ideal process, conditioned by a partial understanding of our cognitive limitations. Our reasons for what we end up doing are therefore only "undefeated" by competing reasons for those alternatives we consider. If someone completely lacked the capacity for technical reasoning, but was presented with a large selection of means-end bundles, she could make the same decision, or a better one, than we do by considering those options "instrumentally".

Inattention to technical priority also makes desire seem more central to practical reasoning than it is. As Vogler (2009, p. 1–4) notes, one concern in ethics is that there might be nothing to practical rationality beyond instrumental rationality. The concern seems to be, "Here is a method by which someone can reason to an end and means using only their skills and desires." This is accurate but misleading, because "using" is too vague. A person reasoning this way who rejects M as an undesirable means to E is not reasoning *about* her desires. She is rather reasoning *with* them, as reading the no-trisection proof is reasoning *with* your desire to do well on your test. All of the actual *reasons* are technical. We could just as well have presented her with all the end-means bundles, from which she could then pick the one she most desires[5]. She then has reason to intend what she chooses only if she has reason to do what she most desires. Any set of criteria that can prioritize complete plans, and function to compare partial plans during efficient decision making, can play the same role.

Here, therefore, is what further practical reasoning must generally "accomplish" given the relative priority of technical rationality, phrased in terms of the theory of technical warrant:

1) It must be able to choose an end among warrantable aims and traims. This amounts to a choice among bundles of potential means-end relations down to the level of basic action,

5. A sophisticated hedonist might choose between complete plans in virtue of additional non-technical reasoning, when some of the options include more than one step. However, this strikes me as less a problem for my argument than a problem for the premise of someone choosing based only on "desire". Deciding between multi-step plans in virtue of reasons likely involves some rational trade-off between, for example, present and future enjoyment. Such trade-offs also provide a foothold for certain ethical views, if a modest one.

given what one knows about one's current circumstances and in light of the technical confidence(s) associated with each bundle[6].

2) It must be able to assist in cutting off those avenues of technical exploration less likely (in some general sense) to yield a better plan, relative to what cognitive resources are available in the time one has to decide.

The simple way of describing this restructuring of further practical rationality is that one only ever has "undefeated" reasons for pursuing a whole plan. To the extent we pursue partial plans, we only rationally do so in virtue of a sense of what the complete plan will turn out to be, and the reasons that will favor it. The success condition of instrumental reasoning is not just determining a means to an end, it is arriving at a complete plan that compares favorably to the other complete plans one might arrive at[7].

This conceptual shift does not favor one broad category of ethical theory over another, or at least it does not obviously do so. If one's standards include Deontic rules against murder or lying, those would disallow any potential bundle with a prohibited means, no matter how desirable the end. If the standards are Consequentialist, bundles would be compared only by the relative value of the end, means, and foreseen side effects. And the possibility of *relational* standards within a bundle is, if anything, more straightforward. If one is permitted to *E*, and also to *M*, but not to *M* as a means to *E*, any bundle with that relation would be disallowed. Those bundles that better reflect prudence, courage, or other virtues could be also chosen on that basis.

Finally, the priority of technical rationality does not prevent one from considering as a potential end what one does not know how to do, or even what one suspects is impossible.

6. Such decisions would typically be based in part on predictable side effects of one's basic means, and therefore the likelihood of each means being realized. See Section 11.1.

7. That ethical decisions involve picking the best plan or plans from a set may appear to favor *contrastivism* about further practical reasons. However, technical rationality can no more support contrastive theories than it can any other theory about what one should do. As noted in Snedegar (2017, ch. 2), contrastivism about reasons and contrastivism about justification do not entail one another. And anyway, all that technical rationality requires is that *decisions* are contrastive, which is difficult to doubt. It could be that all of our reasons and justifications for acting are reasons and justifications *simpliciter*, and it is only our need to arrive at a plan without considering every possibility that introduces a contrastive element to our decision-making.

Possible *outcomes* still have different relative values aside from practical considerations. It is good if a child stops drowning because I save her, or because you do, or because a wave pushes her to where she can stand. Nothing stops my considering a valuable outcome as a potential end prior in *time* to any technical consideration. We do not, and probably should not, typically do so because the most valuable outcomes we can conceive of are generally far from our grasp. There may right now be substance that kills a common and deadly form of cancer, which could be manufactured in a chemistry lab and handed off for testing, eventually leading to a great reduction of suffering and premature death. Unfortunately, the existential possibility of such a substance gives no indication of its chemical structure, and therefore the means of manufacturing it. If we typically considered such goals, we would waste our cognitive efforts and fail to arrive at any end.

9.3.1 The "Ought" of Technical Requirements

I have so far argued against transmission without saying much about bootstrapping, but the two are closely related. We see transmission in those cases where our tentative end becomes our actual end, and bootstrapping in theories that try to generalize such cases into a rational principle. The relative priority of technical rationality implies that we ultimately decide between complete plans. It thereby dissolves the problem of bootstrapping (assuming it is not reintroduced somehow by clumsy practical heuristics).

Along the same lines, what makes the use of "ought" in the Simple or Wide-Scope Principles difficult to interpret is that different realms of reasoning have their own "oughts", which these principles conflate. This kind of problem is easier to see in relation to belief, where the distinction between epistemic and practical reasoning is widely acknowledged. If I spend my days multiplying numbers chosen at random, rather than attending to my own needs or those of others, I might have reason for my belief that 3,842 * 7,509 = 28,849,578, in the sense of having epistemic warrant for it, but not have any good reason to have arrived at that belief. Or, if I am childproofing a bedroom, I might have good reason to know the product of 53 and 29, but not have reason for my belief that 53 * 29 = 1533 because I guessed rather than calculated.

The spirit of practical "instrumentality" is captured by the requirement on an aim E that it be technically warranted. On the premise that M is, or is seen as, necessary for E, M must occur in E's justification. E "ought" to be technically warranted because to take E as one's aim is to see it as a voluntary truth. To take E as a voluntary truth with improper or missing warrant is like taking P as an involuntary truth with improper or missing warrant—it is a mistake of assessment or measurement[8]. At the practical level, the problem becomes that one decides between complete plans partly on the basis of assessed technical confidence, and inappropriate degrees of confidence lead to inappropriate decisions.

9.4 Technical Agency

The idea that the underlying form of further practical reasoning is not identifying *what* to do, but rather *which* to do, has implications for ethical theorizing. I think it is too early to say much about what would need to change at the level of requirements, or even if general requirements would remain a preferred medium for the modeling of ethical decision-making[9]. All that I can offer now is a third "picture", to contrast with two pictures that I see in the background of debates over ethical skepticism. Everything I say on this subject is speculative, and based less on specific textual evidence than my impression after reading some works of ethical philosophy. At best, it is a starting point.

8. Valaris (2014, §§3–4) approaches something like this conception of the normativity of intention. One difference is that the measure of appropriateness is left implicit in the relation between a norm of acting successfully and a norm of intending so as to meet that norm of acting. It is also true that one should *believe* so that one will meet the norm of, for example, forming further accurate beliefs. Epistemology is in part an investigation of the qualities that facilitate appropriate beliefs, and the present theory is in part an investigation of the qualities that facilitate appropriate intentions.

9. Kolodny (2007) describes the debate, related to but not dependent on the problem of instrumental rationality, over whether requirements that govern *states* or requirements that govern *processes* best characterize human reasoning. To the extent that these requirements are supposed to govern individual steps of reasoning, or the results after individual steps, I suspect both approaches are misguided. Different processes can produce similar final results, and the steps taken in a given process can bear no obvious relation to the rules that govern a final result. Anyone interested in these questions should look into the computer science subject of program correctness proofs, and particularly the role of *invariants* in such proofs.

9.4.1 Questionable Ways of Life

Some ethical arguments make use of a contrast between the decisions of an ethical person, and those of a person without any ethical concerns. Sometimes the purpose is illustrate how the decisions of someone not concerned about ethics would be different from our decisions, isolating the influence of ethics in practical reasoning. Sometimes the contrast is used to argue that anyone lacking ethical principles would necessarily make irrational decisions in some circumstances. And sometimes the argument is that certain non-ethical principles, such as an interest in what happens to oneself not just right now but throughout one's life, have ethical consequences and might therefore entail ethical reasons. Whatever use is made of the two types of agent, the most extreme versions of each of these "people" has become associated with what might be called a "way of life"—a picture of what being an agent of this sort would "be like". The implication is that the ways of life of everyone who is not a saint or psychopath lies somewhere between these extremes.

The picture associated with a non-ethical agent is of someone driven only by *desire*. In its most extreme version, this person only cares about what she most wants *next*, ignoring the consequences for what she might want after that. Her decision is comparable to choosing the next ride at an amusement park. If she wants to go on the same ride again, she can just get back in line. If she instead wants to go on a ride on the other side of the park, that motivates her to construct and act on a plan to get there. That planning—her instrumental reasoning—could change her mind about what ride she most desires. If she determines that a ride is closed, or the time to get there too long, her desire for that goal could be sapped below the level of some other one. Still, it is the relative strength of her desires that structure her decision-making process minute to minute.

The desire picture does not rule out ethical *decision making*, because one might still desire what is ethical and to avoid what isn't. If I want my uncle's money but do not want to kill him, the latter would sap my desire to effect an inheritance. Even so, it does seem to rule out ethical *reasoning* in some sense.

The picture associated with the ethical agent, in contrast, is of someone trying to determine what the ethics of her situation calls for. This amounts to a kind of *calculation*—she applies her ethical principles to what she believes about her situation and determines the best course of action. In this way of life, the significance of choice is mostly symbolic. Once she has calculated, her "choice" will typically be between the best option and other inferior options. The latter may include potential ends that are easier or more desirable. Those might sway us, but they will only tempt her; that she is free to act otherwise just makes her correct choice more commendable. She also faces the problem of *how* to do what the situation calls for, of course, which she solves with instrumental reasoning. If that reasoning fails, she will have to settle for pursuing the next best option. However, generally speaking, it is her situation and ethical principles that drive her decision-making process.

Although these two pictures are quite different in other ways, technical rationality plays a similar role in each. Arriving at a means is "instrumental" because the end comes first, not just in time but in priority. The focus of the instrumental principles I discuss in Section 9.1 also reflect this way of thinking. When there are multiple means to a given end, instrumental reasoning involves further decisions, which are difficult to specify in a concise requirement. On this view, it is convenient, and no great explanatory loss, to focus on on those cases in which there is only one means. If your situation calls for your end to be *E*, and your only available means to that end is *M*ing, then your situation calls for you to *M*.

Neither of these pictures is at all plausible as I have described them, for reasons that I have already touched on. At a given time, a person may desire many things that are totally beyond her grasp more strongly than anything she can currently do. Trying and failing to arrive at means to each of these would lead primarily to inaction. And there are typically all sorts of ways that one could act to better the world, over a short or a long term, if one only knew how. Ethicists are, of course, aware of these facts, and the need for a conception of practical reasoning to accommodate them. However, the usual solution, as far as I can tell, is to adjust the pictures in virtue of the general principle of *ought implies can*: An agent decides what to do in virtue of her desires or ethical principles, given her situation and *implicitly discarding*

whatever she cannot do.

9.4.2 The Neglected Rationality of *Ought Implies Might*

This adjustment is not unworkable or incoherent. It does imply a rather different way of life, however, in ways that seem to have gone largely unnoticed. The first clue that something is off is in how it alters the role of instrumental reasoning. There is now a reason for why it is so often successful—it should presumably *always* yield a means because it is just the working out of what one in some sense already knows. This is very pessimistic. When you greatly desire something and are not sure whether you can attain it, you will often think about how you might. And if someone is nearby and in danger, you can and should think about how you might help, even if you start out with no idea of how to do so. In the adjusted pictures, the desire-driven practical reasoner is technically lazy—she never pursues any desire she isn't already confident she can realize. And the ethics-driven reasoner is a scientist when it comes to further practical problems, but a dilettante when it comes to technical ones.

One of the most striking aspects of the recent debates about instrumental rationality is how little attention has been given to the problem of finding that one has no means to a potential end. While "ought implies can" is just a slogan, it might be translated into a principle like the following:

> (No Means Principle): If you believe that you have no available means to *E*, it is not the case that you ought to *E*.

Like the Simple Principle, this seems like a straightforward statement of an obvious requirement. And similar to that principle, it is vulnerable to "negative bootstrapping". If you text me asking for help moving your couch as my ferry is leaving the terminal, the fact that I cannot be back in time gets me off the hook. But if I get your text when I only intend to catch the ferry, my taking it anyway does not just erase the obligation. One way to circumvent that problem would be to make the condition "if you believe *now*", so that the principle does not apply when you still have available means. Unless one is obligated to plan right away, however, this new version might still be abused by putting off one's planning. For almost any end *E*, there is

some point in time after which it is too late to *E*. If I know I should *E* but don't want to, why not just put off any planning until it is too late?

There may well be some further formulation of "ought implies can" as a principle of practical reasoning that avoids such problems[10]. But the problem is more difficult for some theories that deny there is any distinct problem of instrumental reasoning. Raz (2005b) and Scanlon (2007) argue that a "facilitating" reason to *M* is, like all other practical reasons, grounded in the value of *M*ing. Such reasons are distinctive from others only in that that value can be located in a distinct, or differently described, end *E*. Every potential *E* of value is, or grounds, a *pro tanto* reason to *M*, if *M* is a means to *E*. This theory is, among other things, an attempt to provide a ground-level for practical reasoning. Whether *E* is your end does not weigh on what reasons are facilitating. And no Simple or Wide-Scope principle is needed to recognize the value of *M*ing.

According to Raz and Scanlon, this unity demonstrates that instrumental reasoning is not a distinctive category. But they do not provide any specific account of why it is irrational to intend to *E* when one has no available means to *E*. Here is one theory that does seem to work, which is equivalent to the pessimistic view: Suppose that there were a creature that could only "see" value in a potential end in virtue of a potential means to that end. That is, it is blind to the value of any event or circumstance that it is not currently aware of some way of bringing about. This creature would have a system of practical reasoning that is both ground-level and unified in the way Raz and Scanlon describe, but it would be very different from us.

We have the capacity to see the value in *E*, and then consider if there is some means *M* to bring it about. But whenever one cannot arrive at a means, one should not take *E* as an

10. I wonder, however, why this literature has put so much faith in independent requirements, as opposed to unified algorithms or some other sort of decision process. I suspect the reason has something to do with a commitment to objective practical reasons, combined with the view that the subjectively right thing to do is ideally or typically determined by those objective reasons one has subjective access to. The idea might be something like: If no complex algorithm is needed to arrive at the objectively best plan, why would one be needed for a subjective decision? Setting this inference aside, I question the premise. Given a starting point of *every* objective *pro tanto* reason, why think that the right thing to do would be straightforward, or determined by independently phrased principles? Or alternatively, in a scheme in which all objective reasons are "objectively undefeated", why think that we ever have any subjective grasp on the *absolutely* best thing to do?

end regardless of the value one sees in it. A ground-level theory must account for this form of consistency at the ground-level. Raz explains necessary-means-end consistency in terms of proper human functioning (Raz 2005b, p. 15). This is not a plausible account of why you should not intend what you have no means to accomplish. Assuming that the value you see in *E* gives you a *pro tanto* reason to *E*[11], not having a means must preclude arriving at an all-things-considered reason to *E*. What you have an all-things-considered reason to do is presumably not only a matter of proper functioning. Some sort of "at least one means" rule is therefore built into the structure of all-things-considered reasons and/or reasoning.

Adding degrees of technical confidence to the overall picture changes it, but not in a way that helps. *Ought* does not really imply *can* so much as *might*[12]. In most situations, the likelihood that one should *E* will increase with the likelihood that one can *E*, if only because there will be other ends that are valuable and that you can intend with more confidence. However, *ought implies might* is not nearly as useful a "filter" for planning. At a given time, a person may desire many things that are just barely within her grasp, and there are typically all sorts of ways one *might* improve one's situation. Even worse, one often has only a very vague sense of technical confidence for a given potential end prior (in time) to carrying out some technical reasoning about it. That is, even if we do have a generally accurate sense of what we might possibly do (which is far from obvious), we do not have a generally accurate sense of our likelihood of doing each of those things[13].

Technical rationality addresses these problems at the ground-level in virtue of technical

11. Otherwise, we're back to the "value blindness" theory.

12. Or, as it were, "Might makes right!"

13. Confronted with the suggestion that "ought implies might", an advocate of "ought implies can" can reply that reasons for an agent to pursue a goal in the face of lower technical confidence are reasons for her to *try*, and she *can* try with high confidence. My likely complaints about this approach may be evident at this point. One objection is that it conflates reasons for pursuing a goal in the face of lower confidence with reasons specifically to try, as in a parent asking her child to "at least try" to dive off the board into the pool. Another is that if we do indeed also try in high confidence cases, then our reasons are presumably *always* reasons to try. But putting things this way seems to leave the actual realization of the goal as a secondary concern. If a child is drowning and I merely make *a little* effort to save her, I have not done what I should have. This is true whether I have high or lower confidence that I can succeed. When all your reasons relate to the goal itself, the fact that you *can try* to realize it should not be a focus of your decision-making.

confidence. Intending to *E* (in a loose sense) means aiming or traiming to *E*. Aiming involves high confidence, and traiming involves some degree of confidence, however vague. One can reason hypothetically about a potential end *E* without being aware of any means to it, but actually traiming to *E* without any idea of how to go about it (or how to go about going about it) is inappropriate, because whatever degree of technical confidence one associates with *E* will be inappropriate. That technical and further practical reasons are separable allows value to be assessed independently of means and technical confidence.

At the same time, however, the relative priority of technical rationality introduces a problem that the pictures of desire- or ethics-driven reasoning do not grapple with: Almost every time you decide on an end and means to that end, you *might* have arrived at a better means or end by reasoning further. Even a small amount of further cognitive effort *could* put a more desirable or ethically favorable end in your grasp. It might also, of course, accomplish nothing. There may no better ends, or none may cross your mind. The domain of what a human agent might do in a given situation is so vast that she can only explore a tiny fraction of it. At the point she decides on an end, she will generally have little reason to think that reasoning a bit longer would not yield a better option.

What one should do, by whatever standard, depends on what one might do, conditioned on how likely it is that one will manage to do it. Lacking that information, which is beyond one's capacity to arrive at, one cannot be sure of what one should do.

9.4.3 The Guise of the Good Enough

As a result, we cannot generally determine *the* right course of action, even by our own standards. Clear cut situations, such as those ethical thought experiments that stipulate only two or three ends as relevant, are possible but rare. In an actual trolley scenario, you would likely pull the lever (or not), or push the fat man (or not), at the last possible second, after desperately searching for any better option. Indeed, part of what makes such cases alien to experience is the (often implicit) idea that everyone is or will be in agreement about the technical possibilities. It is one thing to be haunted for the remainder of your life for causing one death

necessary to save many lives. It is another to find out shortly after acting that you could have short-circuited the terminals in the electrical box a few steps away, shutting down power to the trolley and saving everyone's life.

Generally speaking, deciding on what to do, and therefore what to intend, is in part deciding that you have reasoned *enough*, given your circumstances. This "settling" on a course of action is most evident when one makes the least effort. When I just grab a bag of potato chips and crawl back into bed, my excuse for this "lunch" may not be that other options seem like too much bother, but that even *considering* other options seems like too much bother. But even a saint could not explore every relevant possibility before acting[14]. A very praiseworthy plan might still be improved, and whoever undertakes that plan will typically not be sure she could not have improved on it. As the saying goes, "perfection is the enemy of the good". The typical goal of decision-making is not a superlative end. It is a bundle of end, means, and foreseeable consequences that strikes a good balance between the strengths of one's desires and ethical commitments, one's knowledge of the relevant circumstances and potentially competing goals, and the cognitive effort of the decision-making itself. We generally act under the guise of the *good enough*.

To be disappointed that we cannot often knowingly act in *the* right way is, I suspect, to feel the loss of some version of the calculation picture. As I have argued, however, that picture is ultimately pessimistic. Only a creature that can only act basically, or that has a very limited capacity for symbolic reasoning, could generally act in ideal correspondence with its own norms. We cannot do so because of the great extent of our potential ends, as measured by what outcomes we might technically justify at the limits of our capacity for reason. This is a good problem to have. The understandable focus in ethics on cases that limit technical possibilities should not obscure the great benefit of our capacity to warrant novel, situation-appropriate ends that we might never have even contemplated in any other circumstance.

14. At least, a saint at large in the world cannot. There are certainly religious ways of life in which ends are largely determined by tradition or doctrine. I do not mean to dismiss such practices as less ethical. Still, they largely obviate decision-making itself, except in the sense of deciding to continue down that path as opposed to another one, and therefore lie outside of the ethical debates I am considering in this chapter.

What I can offer, at this early stage, is a third picture of agency to represent an ideal of practical problem-solving in its open-ended form[15] Even if we cannot often approach ethical challenges like a scientist, we can still approach them like an engineer. The technical justification of a non-basic aim is a constructive process, in that one chooses "premises" to secure one's "conclusion". The most exemplary technical reasoning is *creative* in both senses of that word. When an agent, in virtue of her skill and knowledge of the situation, rips the cover off the electrical box, shorts the circuit, and thereby stops the trolley, she acts well not only technically but also ethically. An engineer rarely has reason to think she has arrived at the *best* solution. Still, the best engineering demonstrates just how excellently "good enough" a solution can be.

The fact that one generally has to settle on, and therefore for, one's end also means that even the most ethically minded person must make substantive choices. The basis of one's decision to stop considering alternatives and intend to E is so difficult to characterize in part because of how many realms of cognition it draws on. It should surprise us when different people tend to make the same decision in a given circumstance, rather than when they do not. For these reasons, the priority and extent of human technical rationality implies something in the neighborhood of *authorship*, a view that is sometimes seen as in conflict with instrumental rationality (Millgram 2016). The ends that you pursue are yours in virtue of how many aspects of yourself you draw on in deciding on them[16].

15. If the subject were just decision-making in isolation, the most open-ended form of that is probably an artist's way of life. However, for the picture to serve the needs of ethical theorizing, it must be more directly applicable to ethical problems in the moment.

16. This kind of authorship different from that of Velleman (1992), given that an agent's conflicted decisions are just as much decisions from a technical perspective. However, it does appear to be compatible with his theory.

Korsgaard's earlier views on authorship (Korsgaard 2008) seem more difficult to reconcile with technical priority. One problem is that she makes a distinction between actions, such as walking in order to get to the store, and (mere) acts, such as walking (p. 219). Given that the only reason inherent to the former is a means-end reason, from a technical perspective these are just two complete plans, which should be acted on or not for one's further practical reasons. In more recent work, however, her views have evolved in a direction more compatible with, and even suggestive of, the priority of technical rationality (Korsgaard 2009, ch. 4).

The guise of the good enough also at least complicates any claim to Kantian universality. I think it is correct that when someone settles on an end, she should be willing to say that if anyone else were "in her shoes", that person could settle on the same end. But this is "universal" only in a very weak sense.

9.4.4 The Good, the Doable, and the Right

In the abstract, the priority of technical rationality does not support one ethical "approach" over another because each of those approaches incorporates some variant of *ought implies can*—usually as one requirement on equal footing with many others. Any theory that incorporates such a principle is already grappling with something close to priority, so it is plausible that any further adjustments would be formal rather than substantial. Even so, the tentative arguments of this chapter might have substantial ethical implications. If what one should do depends on what one can do, then general patterns of good action can depend on general patterns of potential action, which differ not just from person to person, but between social groups and across time.

This possibility is easiest to illustrate with a reductive example.

The domain of potential high-confidence ends of a typical person in ancient Greece was extremely limited compared to ours. The domain of lower confidence possibilities was also surely smaller, and the respective degrees of appropriate confidence generally lower. Roman society was more complex, and that complexity included new technologies and institutions that increased the range of potential actions. There was more that people could confidently do, and more still in the case of those whose writing has survived. Even so, what learned Europeans then, and through the middle ages, were able to accurately predict about the consequences of their actions was modest. Later, in the Enlightenment and, perhaps more importantly, the industrial revolution, the extent of what someone might reliably do—what aims a person could technically justify—exploded.

These three cultural eras are also associated with Virtue Ethical, Deontological, and Consequentialist frameworks, respectively, in that some of the first descriptions of each were written by philosophers living in the corresponding era. Contemporary ethicists who favor a given framework usually argue for them in ways that apply to any culture or time. Given the priority of technical rationality, however, it is also possible that each system is not just appropriate, but *most* appropriate, to its era. This proposal does not entail any form of relativism. The

people of each cultural era could still have roughly the same sense of what is *good*—and to the extent they disagree, some could be right and the others wrong. Theological arguments aside, the Lisbon earthquake was *bad*, rather than morally *wrong*, and everyone confronted with the facts could see it that way. The further question of what is *right* involves a kind of arbitration between what is good and what one can do. General trends in the domain of potential ends might therefore call for different standards for ethical decisions. As the extent of what people can do increases, a method of arbitration that had worked well before may start to fail.

What I am proposing is that the part of ethics that extends our conception of what is good to our conception of what is right—assuming, controversially, that there is such a part—may itself be a sort of technique. And as with any technique, on learning something new it may become obsolete. I have focused on European theories in this example, but other cultures and eras have their own skills, technologies, and institutions, and these could make other frameworks appropriate, also without any relativism. Also, if there is such a dependence, it is also worth noting that simply gaining too many techniques all at once could lead to moral disaster without anyone (necessarily) being at fault, if it is the culture's conception of *at fault* that ceases to function well.

These illustrations are almost certainly far too reductive. It would be very surprising, for example, if a general shift towards deontology did not preserve some role for virtue ethics. At the same time, there are today not just ends but *accidents* that can cause more suffering and death than virtually anyone in ancient Greece could hope to intentionally bring about in her lifetime, if she so wished. We should be at least as surprised, I think, if the same framework were appropriate to every ethical context.

CHAPTER 10

Intentional Action

The theory of technical warrant falls just short of a theory of intentional action. If you consistently meet the normative requirements of aim and traim, you will generally have an accurate sense of the likelihood of bringing about your goal. And the nature of those attitudes make it likely that you will recall your goal when it is time to act. The rest of your "participation" in a basic action consists of an exercise of your will—which should pose no challenge when your further practical reasons are in order—and your subsequent attempt to realize the basic aim or traim. The realization of one's non-basic aims will generally follow both temporally and, in virtue of technical justification, rationally.

Still, I have so far said little about action itself, or intentional responsibility, or any other aspect of the larger subject that is not related in one way or another to prospective confidence and planning. The motivation for the theory of the previous chapters was the possibility of accounting for appropriate technical confidence on the model of credence and epistemic warrant. In one sense, that thought can take us no further. Any theory of intentional action also has to do with what happens when the agent acts, or attempts to act. In another sense, however, the theory of this chapter is just another step along the same path. I simply extend the model of Chapter 5 to include knowledge itself, and further adapt it to account for intentional responsibility.

That path may appear to be blocked by a large and immovable obstacle. How can the theory be extended to behavior itself—what Anscombe calls *performance*—without their being some aspect of epistemology to correspond to it? And what could that component be, given

that one's epistemic "work" is generally done once a belief or suspicion is arrived at[1], which already corresponds in the technical realm to the warranting of aims and traims? As it happens, however, no such correspondence is needed. Behavior is already to related its (potentially) corresponding basic aim or traim in virtue of entitlement, and to further non-basic aims in virtue of the "glue" of justification. Actions are the subject matter of technical attitudes, as more general facts are the subject matter of epistemic attitudes. All that is needed to extend the theory is further theoretical understanding of that relation.

As I discuss in Section 1.2.1, Anscombe's example of the shopper and the detective illustrates the fact that knowledge and intentional action are both a kind of success, attributable to the knower or actor. The detective has knowledge in virtue of what she perceives and her reasoning, while the shopper is responsible for bringing about certain circumstances—obtaining the items on his list—in virtue of his reasoning and behavior. The same can be said of Obadiah and Agatha respectively, when things have gone, or go, as each expects. When things do *not* go as expected, a belief usually turns out to be false, and an intention usually goes unrealized. But true beliefs and realized intentions are also compatible with "failure" in this sense. Such cases are discussed in the epistemology literature as "Gettier problems", while the corresponding problems of intention have been called "wayward" or "deviant" in the context of Causalist theories.

At the root of each Gettier problem is a discrepancy between the warrant for a belief and circumstances in the world. Aims are subject to analogous norms, and technical and epistemic justifications share the same patterns of reasoning. That commonality of warrant suggests a corresponding commonality of invalidating circumstances. That is, it suggests that the circumstances in a Gettier problem, or technical analogues of those circumstances, would also invalidate a technical justification of the same pattern. And it suggests that circumstances in which an aim is unintentionally realized, or epistemic analogues of those circumstances, would also prevent a true belief with the same content and corresponding justification from counting as knowledge.

1. Other than the potential need to sort out inconsistencies with other beliefs.

In this chapter I argue that because Gettier and waywardness circumstances do correspond in this way, they are really the same class of problem exhibited in different but closely analogous realms of reasoning. This is evidence that the similar-but-opposite relationship between the two is deeper than a need to *somehow* provide for appropriate confidence. If our capacity for appropriate technical confidence were based on different principles and mechanisms than our capacity for appropriate credence, we would expect each to succeed and fail in different ways. Instead, everything needed to complete the "inverse theory" of intentional action is also the analogue of a requirement in epistemology.

10.1 Truth

The first remaining requirement on knowledge is straightforward and uncontroversial. It is almost a universal view among contemporary analytic philosophers that to be knowledge, a belief must be true—you don't know what isn't so. Adapting this criterion to the technical realm leaves it little changed in form, spirit, or accuracy. For an aim to be something one intentionally does, it must be realized—you don't do what doesn't happen. Not every theory of action treats intentional content as propositional, and therefore as something that can be true. More generally, however, all theories do seem to agree that for there to be an intentional action, what was intended must come about. Given the further modest assumption that intending entails *representing* one's goal, its coming about is equivalent to its representation becoming accurate, to at least some degree.

The flip-side of the looser requirements on aim—compared with, for example, Cognitivist theories of intention—is that the truth of a technical attitude rules out fewer non-intentional possibilities. The only requirement on the content of an aim is that it be a proposition[2], and therefore an aim can be realized without any involvement on the part of the one who aims. If Obadiah makes the peanut butter sandwich, then Agatha has certainly not made it. Nevertheless, her aim that there will be a peanut butter sandwich turns out to be true. Any theory a

2. See Section 6.1.

non-metacognitive "ground-level" of technical representation faces the same difficulty.

The larger theory of intentional action therefore needs further requirements to mark such cases as unintentional. It turns out, however, that it does not need further *specialized* requirements. The framework that addresses the hardest cases—those that self-referential or causation truth conditions fail to rule out—also addresses these easier ones. I therefore put off that discussion for Section 10.2.4.

10.2 "Waywardness"

I have already discussed some theories of what makes the realization of a goal intentional. Frankfurt, for example, argues for a condition of *guidance* of the action on the part of the agent[3]. His position is close to Anscombe's, who argues that as long as you have prepared properly, and don't change your mind, your failure to act successfully is a failure of your "performance". However, as I note in Section 3.1, the dominant view in recent action theory is Causalism, according to which an intention—considered as the physical instantiation of a mental state—must be an efficient cause of the behavior leading to its realization. The content added to Cognitivist intention-beliefs is sometimes designed to "enforce" some variant of Causalism. There is general agreement, however, that any or all of these additional requirements can hold in a case where the corresponding action is not intentional. The problematic cases are most often discussed in relation to Causalism under the heading "causal waywardness" or "causal deviance", but some also cause problems for other theories. I will therefore refer to all such cases as simply "wayward".

Since the 1960's, the dominant theory of knowledge in analytic philosophy has been that it is *(epistemically) justified true belief* with an absence of so-called *Gettier problems*. Strictly speaking, a Gettier problem is sort of gap between the *justification* for a belief and circumstances in the world, which prevent the belief from counting as knowledge even though it is true. However, the generalization of justification into *warrant* brings with it analogous gaps

3. See Section 3.5.

between the *entitlement* for a belief and circumstances in world. In the case of a perceptual belief, the relevant circumstances usually occur in the part of the "world" constituted by the believer's perceptual system. By the 1980's, Gettier problems had already been so thoroughly discussed that there are already expressions of frustration and fatigue in papers of that time. And there is still no theory of problem-precluding conditions that is both complete and widely-held. This and other factors have led some to reject the Gettier framework entirely, a position most strongly associated with Timothy Williamson.

Setting that fatigue aside for Section 10.3, and temporarily limiting the discussion to cases of high confidence, the corresponding technical theory would be that an intentional action is a *(technically) warranted true aim* with an absence of … a certain kind of problem. In the case of a justified aim, the problem would be a gap between the justification and circumstances in the world, which prevent the aim from counting as intentional even though it is true. And in the case of aim that carries an entitlement, it would be a gap between the entitlement and worldly circumstances that, generally speaking, occur in the part of the "world" constituted by the agent's "motor control" system.

The central theory of this chapter, and the keystone of the thesis as a whole, is that *given the theory of aim, traim, and technical warrant*, each variety of waywardness corresponds either to a Gettier problem or a corresponding problem of epistemic entitlement. With respect to justification, the correspondence can be stated precisely: Consider any combination of technical justification and circumstances in the world above the level of entitlement that prevent the justified true aim from being the agent's intentional action. If a belief with the same content as the aim were epistemically justified in virtue of the same pattern of reasoning, it would be true but fail to qualify as knowledge. As I discuss in Section 10.2.6, the reverse is not always true, or at least there is some evidence for counter-examples. Otherwise, however, the correspondence is exact, and provides further evidence for the inverse relationship between knowledge and intentional action.

That Gettier problems and waywardness are somehow related is not a new idea. The association is probably most often mentioned in passing as in Levy (2013, p. 709), but some theories

have made substantial use of it. Gibbons (2001) argues that we might explain intentional action in terms of our knowledge of what we do, and that Gettier problems are therefore revealing of both. His approach is along the same lines as Cognitivism and Inferentialism, in that it is an attempt to "import" elements of epistemology directly into action theory without adapting their direction of fit. Timothy Williamson takes up the parallel indirectly in the context of his "knowledge-first" program, while rejecting the reductive account of knowledge that admits of Gettier problems. I discuss his view in Section 10.3.1. Yair Levy also uses Williamson's epistemic framework as a starting point for his "Intentional Action First" theory (Levy 2013), but less as a theory that he adapts than as a source of objections to existing reductive approaches. Finally, Gettier problems are a theme in recent "Intellectualist" debates over whether "knowledge how" is a form of "knowledge that" (Cath 2015).

Unfortunately, these theories are limited by how they draw the parallel through the concept *accident*. "Gettiered" beliefs are accidentally true, and in a case of waywardness an agent achieves her intended goal accidentally. Each is an unusual sort of accident, in that the result is as it should have been. However, this is often seen as simply a *coincidence*, which is not much more specific. There are all kinds of accidents and coincidences, and picking out a phenomenon with one of these concepts does not reveal much about it. The question "What specific relations of end-intention to means-intention would allow the former to be accidentally realized even when the latter is intentionally realized?" could be a fruitful line of investigation, which might lead to something like the analysis of Chapter 7. However, that or similar questions do not seem to have had much influence on theories of intention.

In any case, to determine what has or has not been done intentionally, one must take into account not just the relation of end to world, or means to world, or end to means, but all of these together. Conceiving of intention as a technically warranted aim captures the full relations of end to ultimate means. The established "categories" of waywardness can accordingly be distinguished and explained in relation to what element of technical warrant is the source of the problem. I now turn to those categories, again starting with entitlement and then moving to justification.

10.2.1 Exceptions for True Aims with Entitlement

It has been argued that one can be epistemically entitled to a true belief without it being something that one knows. The most widely discussed (hypothetical) example is a perceptual belief derived from an accurate or "veridical" hallucination.

A hallucination—that you are not *aware* is a hallucination—can entitle you to false beliefs about your immediate environment. In virtue of the abnormal functioning of my visual system, I could have a visual experience of a pink rabbit. Barring self-awareness that my vision is not reliable, or further reasoning about the likelihood of there being a pink rabbit in my room, I am as entitled to believe that there is a pink rabbit as I am to any other visual belief, given that I have no way of directly distinguishing the hallucinatory experience from a normal one.

It would also be possible, if very unlikely, for someone to hallucinate what is actually happening. Hallucinations are the result of abnormal functioning of a perceptual system, and an abnormally produced visual experience is a hallucination regardless of its content. Lewis (1980, p. 242) and others have argued that any perceptual belief derived from a veridical hallucination would not constitute knowledge even if it is true[4]. Hallucinatory content is not produced by a reliable process, and it is not explained by the past functioning of the perceptual system through evolutionary history.

Another related but distinct case is that of a "perceptual belief" that is not actually derived from any perceptual experience. Suppose that I believe that my car is downstairs, and if you were to ask me why I would say "Because I just saw it there." I might offer this explanation even if I am mistaken about having seen my car, either because I have a false memory of the visual experience, or because the belief itself was produced by abnormal cognition. Because I cannot distinguish the belief from one that was derived from experience, or could only distinguish it through reflection that I have no reason to pursue, I am similarly entitled to it.

4. Similar cases are sometimes discussed under the "Gettier" heading, as when a very realistic hologram is mistaken for a vase that is nevertheless nearby (Lehrer and Paxson 1969, pp. 234–5). What distinguishes these cases from Lewis's is the evidence one needs to work around them—something about the environment in the case of a hologram, versus something about one's own capacities in the case of a hallucination.

The analogous problems of technical entitlement are cases of "antecedential" (or "primary") waywardness. The term "antecedential" comes from discussion of the examples in relation to Causalist theories. Supposing that a causal chain stretches from an intention, through some behavior, to the realization of the goal, antecedential waywardness is so named because it occurs in the part of the chain *before* the behavior. A classic case comes from Davidson (1973, p. 79):

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally.

In terms of technical warrant, the climber's end is his aim that he no longer suffer the weight and danger of holding another man on a rope. Given some fairly simple reasoning from background knowledge, he can technically justify that aim in part by an aim to let go of the rope. The act of letting go is about as basic as action gets, and therefore that aim would carry a technical entitlement, making both it and the aim it justifies properly warranted. There are then two ways of filling in the example.

In the more likely version, the nervousness alone makes the climber let go of the rope before any decision to act or exercise of will. Although I have left the specifics of will or volition an open question, the technical entitlement to a basic aim depends in part on an exercise of will in awareness of the aim and its condition having been met. Willing connects the aim to the source of its entitlement, and without that connection, confidence in *that* entitlement cannot be vindicated by *that* realization. This version of the example is analogous to a perceptual belief that is not connected to any perceptual experience.

In the other version of the example, the climber does initiate letting go, but his nervousness interferes with his motor control. Suppose, for example, that even if his aim had been to grab tighter, he still would have let go due to that interference. The poor functioning of his capacity for gripping is therefore inconsistent with his entitlement to a reliable exercise of that capacity, even though his attempt to let go happened to have the desired result. This version is analogous

to a perceptual belief derived from a veridical hallucination[5].

In either version, the climber's ridding himself of the weight and danger is also unintentional, because that end is justified in part by the compromised basic aim. If I were to justify a belief that my car is less than two blocks away by my compromised belief that I just saw it outside, I would not know that it is two blocks away even if it is. If my friend borrowed the car without asking, and only dropped it off a block in the other direction two minutes ago, I have no good evidence of its current location. In both the epistemic and technical realms, compromised warrant is transmitted through patterns of justification.

10.2.2 Exceptions for Justified True Aims

When the issue with the epistemic justification of a true belief does not ultimately trace to an entitlement, it is called a "Gettier problem", after Edmund Gettier's influential paper on the subject (Gettier 1963). It is difficult to generalize about Gettier problems as a class, but here is a simplified version of the second example from his paper: Smith has a justified belief that Jones owns a Ford, which justifies his further belief that either Jones owns a Ford or Brown is in Boston. Smith has no evidence at all of Brown's whereabouts, but his evidence about Jones's car provides solid support for his disjunctive belief. As it happens, Jones doesn't own a Ford—perhaps the car Smith knows of is borrowed and Jones lied about it—but Brown is coincidentally in Boston. So Smith's disjunctive belief is epistemically justified and true, but does not constitute knowledge, because what evidence he has does not connect in the right way with how the world is.

However difficult it is to generalize about Gettier problems, they can be systematically linked to the remaining two types of waywardness by the patterns of reasoning common to epistemic and technical justification. In almost all cases, a gap between justification and cir-

5. Burge, following Searle, argues that there is no such thing as veridical hallucination, because perceptual content includes a minimal reference to the system's proper functioning (Burge 2005, pp. 25–6). If the system is not functioning well, the experience cannot be fully veridical given this reference. However, this view is just as easily adapted to the technical realm. Given that the object of a basic aim is to move one's body, it is if anything more intuitive that such aims could include a kind of reference to the proper functioning of one's "motor control".

cumstances that prevents an epistemically justified true belief from being knowledge will also prevent a technically justified true aim from being an intentional action. The link is easiest to explore by adapting a case to include Agatha and Obadiah, and considering whether Agatha acts intentionally with respect to some aim in relation and whether Obadiah's belief constitutes knowledge. Consider, for example, this version of Chisholm's avunculicide case, modified to more clearly support high confidence and to have Obadiah along for the ride[6]:

> Agatha wants to inherit her uncle's money, and so she plans to kill him by putting a motion-sensitive bomb in his house. She has good reason to think that he will be away for several hours in the evening but will then return. On the way to his house, her car hits another, which runs into a wall and catches on fire. Having a bomb in her car, Agatha drives to her Uncle's house without stopping, where she places the bomb and leaves. As always, Obadiah is with her and observes all of these events. The other car was coincidentally driven by her uncle, who died in the fire.

In the action theory literature, this case falls under the heading of "consequential" (or "secondary") waywardness. As antecedential waywardness is said to be located in the chain between the intention and the behavior, consequential waywardness is, on the Causalist analysis, located between the behavior and the realization of the intention. However, given that Agatha completes each individual step of her plan to kill her uncle, the case also poses a challenge for theories of guidance or performance.

Making certain assumptions—such as the availability of reliable cars and motion-sensitive bombs, and Agatha's knowledge of her uncle's last testament and schedule—Agatha's aim to inherit her uncle's money could support high confidence. It would be technically justified in part by her aim to plant the bomb, which would in turn be justified by her aim to drive to his house. If Agatha shares her plan with Obadiah, and he is confident that she can realize her basic means, he could justify a belief that she will inherit the money using the same reasoning she does. Or, the next day, he could justify that belief in virtue of some of that reasoning and his having witnessed her place the bomb. Both her aim and his belief would then be justified

6. The many examples in this lineage are always credited to Chisholm (1966), but I have not confirmed the reference.

and true, but she would not inherit the money intentionally, and he would not know that she will inherit it[7].

Problems of "tertiary" waywardness—a sort of catch-all for cases in which the relevant problem cannot be traced to a point on a causal chain—can be evaluated in the same way. Consider this example from Mele (1987, p. 56):

> Fred is taking a machine-readable multiple choice test. His strategy is to circle on the question sheet the identifying letters to the answers that he feels certain are correct and then, after all such circling is completed, to fill in the corresponding spaces on his answer sheet. At this point, he will take up the more difficult questions.
>
> An hour has elapsed, and Fred is reading the forty-fifth question. He is confident that the answer is 'bee', which word appears next to the letter 'a' on his question sheet. However, as a result of of an understandable slip of the pencil, he circles the letter 'b'. As luck would have it, 'b' is the correct answer. Later, when filling in the answer sheet, Fred looks at the circled 'b' under question 45 and fills in the space under 'b' on his answer sheet—intending thereby to provide the right answer.

The focus of this case is on the time when Fred fills in his answer to question 45. To adapt it for consistently high confidence, we can suppose that he remembers that question by number and his confidence at the time he recorded his answer. When he gets to the space on the answer sheet for question 45, Fred's aim is to fill in the right answer. That aim is technically justified in part by an aim to copy the answer from the question sheet, which may be basic or broken down into steps, but which for we can assume he does with confidence and intentionally. His aim to answer correctly is also technically justified in part by his belief that he recorded the right answer, which is in turn justified by his grasp of its subject matter.

However, because that "grasp" should have resulted in his recording the wrong answer "a", his belief that he recorded the right answer suffers from a Gettier problem. If Obadiah were looking on when the answer was being copied[8], his belief that the correct answer was recorded would therefore also suffer a Gettier problem. As I note in the last section, problems of warrant are transmitted through patterns of justification. In Obadiah's case, it will be transmitted to his

7. At least according to the familiar analysis. See Section 10.3.

8. Assuming that he has the sort of insight into Fred's skills and beliefs as he has into Agatha's.

belief that Fred filled in the right space on the sheet for question 45. That belief corresponds to Fred's aim to fill in the right answer. For Fred, what starts as a Gettier problem becomes waywardness as it is transmitted into the technical realm, making the realization of his aim unintentional.

In relation to the theory of technical justification, therefore, what distinguishes secondary and tertiary waywardness is not a difference of relative position in a causal chain, but a difference of attitude towards the proposition at the root of the problem. Those cases that are called "consequential" trace most directly to a justifying aim, while those that are called "tertiary" trace to a justifying belief. In either case, however, the problem is ultimately a gap in justification, and can be analyzed the same way.

10.2.3 Exceptions for True Beliefs with Entitlement

This explanation of ultimate epistemic and technical "success" suggests a fourth category of waywardness as commonly understood. Cases in this class would probably be labeled "tertiary" by the Causalist standard, but they do not seem to have been much discussed in the literature, or at least I have not turned up papers that discuss them. Regardless, if secondary and tertiary waywardness are problems of justification, but differ by type of justifying attitude, the same distinction should apply to entitlement. Suppose, for example, that Agatha decides to simply shoot her uncle, and does so while having a veridical hallucination of him. Although her perceptual belief that her uncle is in front of her is true, it is not knowledge, because it is derived from an hallucination. By the same analysis as a tertiary case, the realization of any aim justified by that belief would not be intentional.

10.2.4 Non-causal Realization

The unification of Gettier problems and waywardness is the culmination of the metaphorical inverse relation between knowledge and intentional action. In both realms, the requirement beyond truth and warrant is not the existence of a causal-in-the-right-way chain or of good

guidance or performance, but that the warrant bear the right relation to relevant worldly circumstances. The realization of an aim is intentional or accidental in relation to its warrant, but *accidental* contrasts with *intentional* specifically because the attitude of aim assesses voluntary truth. In the case of non-basic action, technical justification provides a three-way relation between end, means, and circumstances to account for both appropriate confidence and intentional responsibility.

The same relation can account for cases in which there is no, or no relevant, causal chain from the agent to an aim's realization. Agatha is not intentionally responsible for a sandwich made by Obadiah because her aim to make it is justified by her basic aims. If, in the previous moment, Obadiah believed that a sandwich would result from her basic actions, that justification for his belief would not give it the status of knowledge once he takes over. And if Obadiah were instead to grab her hands and perform the tasks with them, the result would be like a case of antecedential waywardness, except with a cause external to her body.

Alternatively, suppose that Agatha's plan to inherit her uncle's money were realized as described, except that he is instead killed in an accident with a different, unrelated car. In that case, the nature of Obadiah's corresponding Gettier problem would be basically unchanged. It is true that Agatha's overall situation is different, in that we would not say she killed her uncle accidentally. But that difference is irrelevant to the analysis; she did not kill her uncle intentionally because his death bears the wrong relationship to her reasoning. An agent's responsibility is determined in every case by the relationship between the technical warrant for her aim, and the circumstances of the realization of that aim[9].

9. Although I am largely setting aside the question of partial plans in this thesis, I should briefly clarify how intentional realization of a partial plan is also ruled out. Intentional responsibility for an *explicitly* partial plan can be ruled out by their being no justifying aims or traims by which to judge success. My own, deferred view is that a partial plan depends in part on a positive commitment to plan, which rules out intentional responsibility by any process that does not include that planning.

10.2.5 Lower-confidence Action

The analogy between action and knowledge directly accounts for intentional actions that correspond to aims, but some adjustment is needed for lower confidence traims, as in this example of Daniel Bennett's by way of Davidson (1973, p. 77): "A man may try to kill someone by shooting at him. Suppose the killer misses his victim by a mile, but the shot stampedes a herd of wild pigs that trample the intended victim to death." Here the top-level traim would be to kill, probabilistically justified at low but significant confidence by a traim to shoot the victim, together with knowledge of the damage done by bullets and other background beliefs.

If the man's skill at shooting were excellent, and his miss a freak occurrence, we could consider the example in light of Obadiah's true belief that the victim *will* die, justified in part by knowledge of the man's skill. That belief would suffer a Gettier problem, and therefore the death would not be picked out as intentional. As it is, Obadiah only has grounds to suspect the man will die. A lower-credence epistemic attitude cannot suffer a proper Gettier problem, because it cannot be the basis of knowledge as normally understood.

However, in "Epistemology Formalized", Sarah Moss argues that such attitudes suffer an analogue of a Gettier problem. One of her examples is Sue's realization, under the following circumstances, that "I probably liked that guy as more than a friend" (Moss 2013, pp. 9–10):

> Sue and her friend Bob enter a psychology study. The study proceeds as follows: each woman is given questionnaires that indicate whether she finds her friend attractive. If she does, she is injected with an anxiety-producing drug before meeting her friend. If she does not, she is injected with a saline solution. Sue is not told about the nature of the experiment. She does probably like Bob as more than a friend. On receiving the anxiety-producing drug and meeting Bob, Sue reflects on her fluttering nerves and raises her credence that she likes Bob as more than a friend.

The attitudes and problems that Moss explores in her paper do seem to describe an epistemic analogue to a true warranted traim[10] that is nevertheless not something the agent does intentionally. Given that Obadiah's suspicion that the man will die is justified by his suspicion that

10. If the lower confidence attitude is towards the proposition "the man dies" it will be true or false depending on whether he dies. So it is not inaccurate to call the traim "true" after the fact.

the man will be shot, the eventual trampling compromises his probabilistic warrant for the former in much the same way.

However, one interesting difference between the epistemic and technical realms is that attitudes of what Moss calls "probabilistic knowledge" have no success condition analogous to conventional belief-based knowledge. A suspicion that assigns a 3 in 4 chance of some event is compatible with that event occurring or its not occurring. Traims, in contrast, do have an explicit success condition: whether one intentionally does what one tries to do. The comparison between cases of trying and probabilistic knowledge may therefore be more illuminating about the latter.

Suppose, for example, that Agatha visits a house that she believes may be her birthplace, based on a note written on the back of an old picture. The picture is actually of a different house, but in a complete coincidence Agatha was born in the unrelated house that she visits. In this case, Agatha's traim to visit her birthplace is realized but it is at least arguable whether she does so intentionally. The most straightforward way to test an intuition about the probabilistic Gettier analogues explored by Moss may therefore be to make a suspicion part of the justification of an otherwise high-confidence traim, and then consider whether the realization of that traim is intentional[11].

One problem not addressed by Moss's theory are the "lower bounds" on warrant for intentionally realizable traims. As I note in Section 2.3.3, although you can try to win the lottery, it is generally accepted that you cannot intentionally win the lottery without, for example, rigging it. The literature on this subject focuses on two criteria. One is just the chance of the event happening, which corresponds in the present theory to the degree of technical confidence. This view is implicit in the focus on winning lotteries—a prototypical low-likelihood event. The other factor is that intentional responsibility depends in part on an exercise of skill. Given that responsibility in high confidence cases does not appear to depend on skill—unless the relevant skill is *warranting with high confidence*—the lower bounds may depend on both factors. Mele and Moser (1994, pp. 56–7) explores the trade-off. I leave this debate as it is.

11. See Mele and Moser (1994, p. 56) for another example.

10.2.6 Different Standards of Epistemic and Technical Success

There is one family of examples in which a pattern of reasoning is considered to confer intentional responsibility when used in a technical justification, but not to yield knowledge in an epistemic one. These are characterized informally as cases in which the agent would describe herself as φing, and *is* φing, but does not *know* that she is φing. Schwenkler (2015, §4.2) offers these variants of Anscombe's pumping example, (part of which is quoted in Section 2.4):

> For example, suppose first that unbeknownst to the man at the pump, the pipe that leads to the water supply is generally quite unreliable, though today the water happens to be flowing through it just fine. Or again, suppose that there are two indistinguishable pumps, one running to the house where the party chiefs live and the other running to a different house, and that the man happens to have noticed only the correct one, and set to work at it.

What distinguishes these examples from others, in my view, is how the agent would integrate the missing information into her reasoning. The man at the pump aims with high confidence to replenish the water supply by pumping. On learning that the pipe is unreliable, he should lower his confidence, but does not otherwise need to change the overall relation of end, means, and beliefs about his circumstances in order to *traim* to replenish the water supply. That traiming would then be successful in a way that confers intentional responsibility. The same is true of the case with two pumps: The pump in front of him has a 50% chance of being connected to the house, so he can traim to replenish the water with it.

The difference in how we assess the cases may therefore trace to the fact that *intentional* is the success condition for both high and lower confidence technical attitudes while *knowledge* is the success condition for high credence attitudes. If facts unknown to the agent compromise the status of the warrant for her aim, but only in a way that would lower her appropriate confidence in it, it makes sense that we still see its realization as intentional. Obadiah could also then rightly *suspect* that she is replenishing the supply. However, that suspicion would not be something he knows, strictly speaking, even though what he suspects is so.

10.3 Gettier Dissidence

The previous discussion of the unification of Gettier problems and waywardness, including possible exceptions, depends on the standard analysis of Gettier problems. That analysis is unquestionably dominant in contemporary epistemology, but there are other views. My primary approach in this chapter therefore reflects a choice, as the epistemic model of Chapter 5 reflects various choices. And as with those other choices, the alternatives do not necessarily count against the larger parallel between knowledge and action, or the theory of technical warrant.

Turri (2016) argues that there is no single unified category of Gettier problems, and instead a number of separable, finer grained ways that evidence can be compromised, which are assessed differently by non-philosophers in his experiment: He presented participants with descriptions of five Gettier and two control scenarios, and asked them to judge whether "Emma" knows that she is leaving the store with a diamond. The "Failed Threat" scenario—"The stone Emma purchases is a diamond. A skilled jewel thief tries to steal it from her pocket before she leaves the store, but he fails"—is judged to be knowledge at a high rate, close to the control in which she simply buys the diamond and leaves with it. The "No Detection Dissimilar Replacement" scenario—"The stone Emma purchases is a fake. A skilled jewel thief tries to steal it from her pocket before she leaves the store, and he succeeds. Long ago, Emma's grandmother secretly sewed a diamond into the pocket of Emma's coat."—is judged to be knowledge at a low rate, close to the control in which she simply buys a fake diamond. The participants judge other Gettier scenarios to be knowledge at statistically significant intermediate rates (pp. 343–4).

Setting aside Turri's *theory*, his *approach* can easily be extended to analogous scenarios of waywardness, providing two distinct kinds of evidence. First, the waywardness data in isolation could be used to develop analogous theories of compromised intentional action. Second, the correlation between Gettier and waywardness data for various Agatha/Obadiah equivalents could be used to explore the link between action and knowledge in a more theory-neutral

way[12]. Results from other experiments already suggest some ways in which assessments of analogous cases are likely to diverge. There may be a "bias" against intentional attribution when an action is considered to be a positive achievement, and for it when an act is considered morally wrong, as compared with epistemic equivalents. Still, those factors and other qualities related to ethics or value could be controlled for by varying them independently across descriptions of different scenarios.

I doubt that the data produced by such experiments would be considered definitive evidence for or against the larger theory of this thesis. Even if the correlation were quite poor, the different assessments might be due to unknown factors rather than a difference in underlying reasoning or mental representation. And if the correlation were quite close, it could still be explained in virtue of some broader concept like *accidental*. Non-definitive evidence is, nevertheless, still evidence.

10.3.1 "Knowledge First"

The "knowledge first" program in epistemology, which is most associated with the philosopher Timothy Williamson, is in part an outright rejection of the Gettier problem analysis of requirements for knowledge. Even so, Williamson himself argues that there is a close relation between knowledge and intentional action, and is similarly inspired by Anscombe's example of the shopper and the detective. And he does recognize a relationship of sorts between Gettier problems and waywardness, in that he takes each as evidence of a misguided approach to the study of its respective subject (Williamson 2017, pp. 10—1)[13]:

> Since belief and truth are necessary but not even jointly sufficient for knowledge, many epistemologists have found it natural to ask: what must be added to belief and truth to make knowledge? Schematically, they seek a solution in 'X' of the equation:
>
> knowledge = belief + truth + X

12. The results would not be entirely theory-neutral, of course, because the premise that such cases *are* equivalent depends in part on the theory of technical warrant.

13. The "attempt" paragraphs in this quote appear in the paper in a different order.

> … The attempt to solve the equation for knowledge generated an extensive but still unsuccessful research programme. Many solutions were proposed, of increasing complexity, but unless circular always eventually succumbed to counterexamples, of increasing complexity, to their necessity or sufficiency.
>
> … Since intention and success are necessary but not even jointly sufficient for action, many philosophers of action have found it natural to ask: what must be added to intention and success to make action? Schematically, they seek a solution in '*Y*' of the equation:
>
> action = intention + success + *Y*
>
> … The attempt to solve the equation for action generated an extensive but still unsuccessful research programme. Many solutions were proposed, of increasing complexity, but unless circular always eventually succumbed to counterexamples, of increasing complexity, to their necessity or sufficiency.

Williamson's "schematic analogy" between knowledge and action is mostly aligned with the theory of this thesis. He also argues that intentions need not represent the agent in their contents, and that grammatical distinctions do not provide strong evidence for a difference in underlying propositional representation (Williamson 2017, pp. 7–8). Most of his discussion focuses on these parallels (p. 5):

knowledge	action
belief	intention
truth	success
falsity	failure
fitting mind to world	fitting world to mind
input to practical reasoning	output from practical reasoning

and on the proper or improper functioning of the cognitive systems underlying these capacities. The combination of improper functioning and accidental success is therefore his alternative to a Gettier problem or waywardness analysis.

Williamson's own theory of the relation between action and knowledge is, however, distinct from his objections to reductive approaches. Other theories are compatible with the latter,

including a slightly modified version of the theory of this thesis. I can only guess that he omits warrant from his analogy because he doubts that it has any equivalent in the realm of action. Most proponents of the "knowledge first" program accept the existence and significance of justification; what they are skeptical of is the prospect for reducing knowledge to belief, truth, justification (or, more broadly, warrant), and other factors.

Whatever Williamson's reasoning, however, his discussion of intentional action does not even include, let alone address, the prospective assessment inherent in intention, which I describe in terms of "appropriate technical confidence". All he tells us about the state of intention is that each represents a goal, which leads to success, or fails to, in virtue of certain processes. There is no suggestion as to how the requirements of Chapter 8 might be satisfied, or how one can appropriately decide between alternative plans in virtue of their chances of success.

The theory of technical rationality fills in that story by taking Anscombe's suggestion of a close relationship between knowledge and intentional action to a further conclusion. And it can accommodate Williamson's arguments against reduction with a simple change to the epistemic model. The corresponding theory of intentional action would still include technical warrant, just as the theory of knowledge includes epistemic warrant. Each would differ from the more standard account primarily in its rejection of Gettier problems and waywardness *as* problems. That change at least alters the status of those problems as evidence, and may reduce it. But the overall unification of the two realms in virtue of what and how they appropriately measure would remain.

CHAPTER 11

Practical Knowledge

The theory of technical justification allows an epistemic attitude to justify a technical one, but I have so far set aside the question of how an aim or traim might warrant a belief, and what content such a belief would have. The capacity to form beliefs about a plan is essential to our ability to provide explanations of what we are doing or trying to do, and also for predicting and evaluating whatever non-intentional consequences of our actions we can foresee. I start this chapter by arguing that the parallel requirements of technical and epistemic warrant can account for an entitlement to derive a belief from one's aim, as long as it is correctly qualified by the possibility of changing one's mind.

The strong association of *knowledge* with *belief* suggests that it is those derived beliefs that would be the proper basis of "knowing what you are doing". I argue, however, that other instances of "practical knowledge" are better explained directly in terms of aims and traims. These include Anscombe's idea of "knowledge without observation", how aims can relate us to future particulars, and Aquinas's conception of practical knowledge as being "the cause of what it understands".

11.1 Epistemic Knowledge of What You Do

To predict the consequences of our actions, and to explain to others what we hope to accomplish, we must somehow derive beliefs from our intentions. This is a broader claim than that we need to form beliefs *about* our intentions. The costs and benefits of reasoning from the belief "I intend to φ" are similar to those that an Inferentialist theory places on reasoning about technical confidence, and any proposal that we typically reason this way faces the same evo-

lutionary objections. On the other hand, one should not simply believe that A on the basis of aiming that A, because doing so would negate one's agency. A belief derived from an aim must somehow encode that its realization is voluntary.

I doubt that there is sufficient cognitive evidence to favor one specific "template" for these beliefs over others. For purposes of illustration, however, I will use "I will intentionally φ unless I change my mind." for the belief derived from an aim to φ[1,2]. This form has the advantage of encoding personal responsibility, while keeping the metacognitive content minimal and separable. The epistemic attitude corresponding to a traim to φ can use the same template, but it will be a *suspicion* that associates a degree of credence to that proposition that corresponds to the degree of technical confidence in the traim.

To evaluate whether an aim makes this belief appropriate, we can first ask whether an "expert" on the subject of technical rationality could justify it. So suppose that Angela and Obadiah have read the earlier chapters of this thesis, and understand both the epistemic model and the theory of technical warrant. And further suppose that Obadiah has a particularly accurate insight into Agatha's mental states, either because of their long relationship, or because we give him some device.

If Agatha has a *basic* aim to φ, both could then justify the belief that *she will φ unless she changes her mind* in virtue of the belief that she has that aim together with the theory of technical entitlement: Agatha's aim to spread the peanut butter is appropriate in virtue of her prospective assessment of high technical confidence, which is partly in virtue of her motor-capacities. And one's having a warranted basic aim typically leads to its intentional realization. Whether Agatha goes ahead with her plan to spread, however, is up to her until she does so, revises her aim, or definitively fails to realize it. Assuming the theory of technical entitlement is correct, the case for high credence in the derived belief is therefore about as good as the

1. To the extent that grammar is relevant to these questions, a propositional aim like "There will be a peanut butter sandwich." would correspond to something like "I will intentionally make a peanut butter sandwich unless I change my mind."

2. "I am intentionally φing." is another option, because intentionally φing now is compatible with changing one's mind later. However, for reasons I discuss in Section 2.3.2, that phrasing is also compatible with trying and failing to φ, which makes it too weak to support the "factive" theory of this section.

case for high confidence in the aim.

Either Agatha or Obadiah could also justify a belief with that template corresponding to her *non-basic aim* that *A*. To warrant the belief that Agatha will make a peanut butter sandwich unless she changes her mind, Obadiah could first form beliefs corresponding to each of her technically justifying basic aims. Then he could evaluate the correctness of the pattern of justification tracing from those aims, and their shared beliefs, to her sandwich aim. If it is valid, Obadiah can conclude that there will be a sandwich unless she changes her mind, and that she will be intentionally responsible for doing so. As with a basic aim, the quality of the epistemic warrant for this belief closely corresponds to the quality of the technical warrant for the non-basic aim that *A*. The same will be true of a traim, either basic or non-basic, and its corresponding suspicion.

What is most important to understand about this imagined process is that what Agatha and Obadiah need to verify about her aims in order to justify the corresponding beliefs is *whether they meet the normative requirements of the attitude of aim.* This is the sort of thinking one might do to verify any accuracy-dependent attitude, such as reassessing one's reasons for a belief. Given the information an agent has, an appropriate aim will always make the belief appropriate, or conversely, a problem with the justification for the belief should undermine the justification for the aim. In other words, while a philosopher who subscribes to the theory of technical warrant can explain why the belief is justified, there is no need for Agatha to do so as long as her aims typically meet technical norms. As I discuss in Section 5.3, she can just be epistemically *entitled* to the belief that corresponds to her aim, and the suspicion that corresponds to her traim.

What this entitlement enables is further *contemplative* or *reflective* reasoning about one's plans, including the consideration of foreseeable side effects. Any further beliefs *directly* justified by it must also be conditioned on not changing one's mind. The justification for any belief or suspicion about what will happen *simpliciter* must therefore also take the strength of one's practical reasons into account. Other factors, such as one's general psychological reliability or strength of will, may also be relevant to appropriate credence about what will happen, to

the extent one is self-aware about such matters[3]. Evolutionary arguments count less against the psychological aspects of *this* reasoning because the capacity to act is independent of it. A creature could have both beliefs and intentions (or "proto-attitudes" corresponding to these) but lack a capacity to derive beliefs *from* its intentions. The ability to *reflect* on a plan may therefore be one part of a more sophisticated suite of metacognitive capacities tracing to a more recent stage of our evolution.

11.1.1 Cognitivism Reconsidered (and Rejected)

The theory that our intentions entitle us to derive certain beliefs is another way of accounting for one of the primary motivations for Cognitivism. The connection between aim and belief that it posits is not necessary; if you have no need to epistemically reflect on your goal, you might never form the belief[4]. Whenever the questions "Will I φ?" or "Am I φing?" are relevant, however, the answer would be available without any effort or reasoning on your part[5]. Even so, a Cognitivist might still object that positing both an aim and a belief derived from it makes

3. To be more specific, I am arguing that it is possible to outright *believe* that one will intentionally φ *in a specific instance.* To justify that belief, an agent would have to reason that the likelihood of a change in her further practical reasons is low. That reasoning would likely depend in part on evidence that a change of circumstances is low. However, because circumstances and further practical reasons often do change before we act, there can be no *entitlement* to such a belief, and one will not often be able to justify it.

4. Paul (2009b, §I) argues against the connection being necessary partly in virtue of proposed counter-examples. Much of her analysis, however, depends on intention not being "a matter of degree": "It does not seem to me that folk psychology admits of intentions coming in degrees, at least not in a way that is correlated with one's expectations of success (rather than, say, one's degree of commitment). The [recently paralyzed] agent's intention to clench his fist seems on the surface of it no weaker for the fact that he is only somewhat confident that it will happen. … If it seems to us that the cognitive commitment can vary in strength and object while the intention remains the same, this is reason to think they are two different attitudes." I do not understand this argument. If the worry is that *attitudes* cannot come in degrees, a suspicion is also a "complete" attitude that encodes a lower degree of credence. Otherwise, folk psychology absolutely admits of *trying* with lower confidence that "is correlated with one's expectations of success", rather than degree of commitment. Even the action theorists who argue for a confidence-neutral interpretation of *intends* usually admit that saying "*X* intends to φ" of a lower confidence case at least "sounds misleading". That "the intention remains the same" is a theoretical conclusion, not a folk-psychological starting point.

5. I take the relation between such beliefs and the question "What am I (intentionally) doing?" to be similar, in that one should not need to *reason* to the answer. However, the phenomenon of "zoning out"—discussion of which I am otherwise deferring—shows that this form of the entitlement is more prone to failure. I am not sure that this should surprise us: "What am I doing?" is in some ways a narrowed form of the question "What are my current intentions, whether or not I am acting on them now?". Generally speaking, the human capacity for "enumerative metacognition" appears to be fairly limited.

the theory unnecessarily complex. Wouldn't it be preferable to combine them into a single epistemic or quasi-epistemic attitude?

In Section 3.2 I consider some modest extensions to more traditional Cognitivist theories and to Setiya's variant. This more radical step of "combining" aim and belief would presumably treat the theory of technical warrant, or something close to it, as the missing theory of epistemic warrant for a Cognitivist intention-belief. Setiya's theory is easier to adapt in this way. When limited to cases of basic action, prospective assessments in virtue of his "knowledge how" could be explained by entitlements to basic "technical beliefs". Because this entitlement would sometimes be in virtue of imagining the scenario, it can account for confidence in specific situations. Non-basic technical beliefs would then be epistemically justified on the model of technical justification. Stating the requirements of this new kind of justification is a bit awkward—"A non-basic technical belief must be justified by at least one technical belief, but otherwise … "—but the difficulties are terminological rather than substantial. Given that technical beliefs would be related their eventual intentional realization by technical warrant, the problematic self-referential content common in Cognitivist theories would have no role, and could be omitted[6]. "Technical suspicions" would substitute for traims, and it could be stipulated that, whatever the level of confidence, these attitudes are non-voluntarily recalled at times appropriate to act on them.

The most obvious problem with this proposal is that it fails to "combine" anything. The requirements for technical belief would be quite different from those of epistemic belief. And permitting epistemic inference from a technical belief that A would permit it to justify an epistemic belief that A. So there must still be a special rule for deriving an epistemic belief from a technical one, to prevent the negation of one's agency. That rule would not be needed if the form of a technical belief were instead "I will intentionally φ unless I change my mind." However, that proposal suffers from the opposite problem, which is that such a belief cannot be the *source* of one's commitment. As with the distinction between the Standard and Wide-

6. We can assume, for the sake of this discussion, that the "desire-like" aspect of these attitudes is enough to distinguish them from a conventional belief with the same content.

Scope Principles of Section 9.1, to encode one's commitment to acting as a belief is to take it as not voluntarily revisable, and to represent one's revisable commitment in a belief is not to encode it[7].

Setiya himself sometimes talks of a desire-like belief and sometimes of an attitude that is desire-like and belief-like. If being "belief-like" amounts to being subject to a standard of accuracy, then on this looser conception we could bring the theory full circle back to the pro-attitudes of aim and traim and call it "Cognitivist". But doing so would be both ahistorical and unhelpful. Actual Cognitivist theories have focused primarily on the seeing-as-true aspect of belief, and only secondarily and half-heartedly on appropriate confidence. All of the typical features of these theories—the abandoning of the warrant requirement, the self-reference, the addition of causal or motivational content—run counter to technical rationality.

11.2 "Practical Knowledge"

When it is properly derived from an aim that *A*, the belief "I will intentionally *A* unless I change my mind." can be something one "knows" in the ordinary, epistemic sense of that term. It is equivalent to the belief "If I do not change my mind, I will intentionally *A*.", which is uninterestingly true given a change of mind. Otherwise, both its truth and its status as knowledge depends on the truth of the aim and the lack of waywardness in its technical warrant[8], with the possible exception of the cases discussed in Section 10.2.6, in which intentional responsibility and knowledge diverge. This belief, rather than the aim, may be the immediate basis of an agent's claim that "I know what I am doing."

We should not just assume, however, that the subject matter of every correct use of "knowledge" or its cognates is a belief, or another broadly epistemic representation. The phrase "knowledge how", for example, most likely refers to broadly technical representations of the

7. This claim is closely related to the discussion of phenomenology in Section 7.4.

8. One could also presumably warrant the simpler belief "I will *A* unless I change my mind", which, changes of mind aside, would be *true* when "I" in any way cause the aim to be realized and *knowledge* when it is realized intentionally.

sort I discuss in Section 6.7. And I argue below that most uses of the term "practical knowledge", with its suggestion of high technical confidence, can be better explicated directly in terms of a warranted aim. With this in mind, I close this thesis by considering a few of the famous questions and claims of *Intention*, focusing less on accurate interpretation of Anscombe, and more on one way of understanding her views in light of technical rationality[9].

11.2.1 Knowledge "Without Observation"

One of the threads running through *Intention* that has been the subject of much subsequent debate is Anscombe's conception of "knowledge without observation". Although her application of this idea to action is sometimes glossed as that "we know without observation what we are doing", she first describes the connection this way: "Now the class of things known without observation is of general interest to our enquiry because the class of intentional actions is a sub-class of it." (Anscombe 1957, §8). One unfortunate aspect of the gloss is that it has focused discussion on times prior to and simultaneous with an agent's attempt to act, whereas Anscombe appears to be at least as interested in our knowledge of what we have *done*. Consider this quote (§29):

> … I have argued that my knowledge of what I do is not by observation. A very clear and interesting case of this is that in which I shut my eyes and write something. I can say what I am writing. And what I say I am writing will almost always in fact appear on the paper. Now here it is clear that my capacity to say what is

9. Here is one satisfying point of convergence that does not rise above the level of a footnote. Anscombe (1957, §31) asks:

> What is the contradictory of a description of one's own intentional action? Is it 'You aren't, in fact'?—E.g. 'You aren't replenishing the house water supply, because the water is running out of a hole in the pipe'? I suggest that it is not. … the contradiction of 'I'm replenishing the house water supply is not [that], but 'Oh, no, you aren't' said by someone who thereupon sets out e.g. to make a hole in the pipe with a pick-axe.

In the epistemic realm, the "contradictory" of a belief is just its negation. The technical analogue would the *aims A* and not-*A*, which directly conform to Anscombe's description. The correspondence even extends to the distinctions of self-representation noted in Section 6.1. If you aim that there be a better book on Kant, I might thwart you by aiming that there not be a better book on Kant. To succeed in that aim, I would have to prevent any such book from being published. If, on the other hand, you aim that *you* write a better book on Kant, I might thwart you by aiming that you not write a better book on Kant. To succeed in *this* aim, I only need to prevent *your* efforts.

> written is not derived from any observation. In practice of course what I write will very likely not go on being very legible if I don't use my eyes; but isn't the role of all our observation-knowledge in knowing what we are doing like the role of the eyes in producing successful writing? That is to say, once given that we have knowledge or opinion about the matter in which we perform intentional actions, our observation is merely and aid, as the eyes are an aid in writing. Someone without eyes may go on writing with a pen that has no more ink in it; or may not realize he is going over the edge of the paper on to the table or overwriting the lines already written; here is where the eyes are useful; but the essential thing he does, namely to write such-and-such, is done without the eyes.

This is, in effect, a description of knowledge being gained *through* the "doing". If, in the example, Anscombe were to keep her eyes shut while leaving the room after she finishes, she would continue to know what she has written. This, she proposes, is also the way in which we have knowledge of what we do even when we are "observing".

With respect to *basic* action, at least, this claim has largely been vindicated by neurological and experimental evidence. The existence of two primary "streams" leading from the visual cortex[10] support a related distinction between *perception* and *observation*. The dorsal stream, leading to areas of the brain associated with spatial representation and motor control, is consistent with visual *perception* serving as an "aid" to basic action. Our concept of *observation*, in contrast, is tied more narrowly to epistemic representation of the sort associated with those areas connected via the ventral stream.

There is also a good deal of evidence that when one acts basically, perceptual experience of the predictable effects of one's behavior is generally *attenuated* (Brown et al. 2013, p. 412). That is, the experience of acting, which centers on broadly technical representation, usually incorporates a faded or de-emphasized perceptual representation of the direct effects of that acting[11]. Interestingly, the attenuation tends to end if and when an attempt goes too badly, suggesting that our motor-control capacities track whether the current circumstances are within, or have fallen outside of, the parameters of the current motor "program".

---

10. See Section 6.4.2.

11. One of the most striking and earliest studied examples of this effect is the fact that one cannot normally tickle oneself.

In cases of high confidence, that tracking can also explain one's sense of success when acting basically. When you toss your crumpled paper into the large recycling bin by the door as you leave, you might look as you do so, or you might not. Because the opening is large and you are close to it, you can look away before you even start to move your hand and arm, and be appropriately confident the paper went into the bin. But if you instead accidentally stumble just before letting go, you might loose your confidence, also without ever looking. This suggests that when acting basically with confidence, one's sense of success is primarily in virtue of broadly technical, rather than epistemic, aspects of experience.

This evidence about the nature of basic action is independent of the theory of technical rationality. What technical justification adds to the picture is an account of our knowledge without observation of what we do, or have done, *non-basically*. While it is true that Anscombe also claims that we do not have such knowledge by "inference"[12], her objection is to the idea that we infer what we do from our *acting*. While the technical justification of an aim that *A* involves something like inference, it is (ultimately) in virtue of a basic *aim* to φ, and must already be in place in order for one's intentionally φing to also count as intentionally *A*ing. Because this justificatory structure is itself technical, rather than epistemic, it can "extend" the non-observational knowledge that one has φed to non-observational knowledge that one has *A*ed.

It is important to distinguish the question of whether knowledge is non-observational in this way from the various ways one can fail to act intentionally, and therefore fail to have such knowledge. Cases of tertiary waywardness, for example, show that not all failures to act are failures of "performance". Suppose that an agent technically justifies an aim that *A* with an aim to φ and a belief that *P*, and then successfully φs, giving her a non-observational sense that she has intentionally *A*ed. If *P* is false or Gettiered, *A* may be false or wayward. But this just means that in this case, her sense turns out to be *wrong*. No one contests that almost every non-basic intention depends in part on beliefs of the agent. Anscombe herself talks of the error in judgment in putting "tackle for catching sharks" on a list for shopping in Oxford. Learning

12. See, for example, Anscombe (1957, §28).

something new might undermine one's confidence in the realization of a past or present aim. Nevertheless, when things have gone or will go well, one has a kind of knowledge through one's past or future acting.

11.2.2 "A Very Queer and Special Sort of Seeing Eye"

Anscombe prefaces her extended discussion of *practical knowledge* with this question (Anscombe 1957, §32):

> Can it be that there is something that modern philosophy has blankly misunderstood: namely what ancient and medieval philosophers meant by *practical knowledge*? Certainly in modern philosophy we have an incorrigibly contemplative conception of knowledge. Knowledge must be something that is judged as such by being in accordance with the facts. The facts, reality, are prior, and dictate what is to be said, if it is knowledge. And this is the explanation of the utter darkness in which we found ourselves. For if there are two knowledges—one by observation, the other in intention—then it looks as if there must be two objects of knowledge; but if one says the objects are the same, one looks hopelessly for the different *mode of contemplative knowledge* in acting, as if there were a very queer and special sort of seeing eye in the middle of acting.

The beliefs to which our aims entitle us could be fairly described as providing a "different *mode of contemplative knowledge* in acting". I am not sure, however, that they qualify as having the same "objects" as the sort of ground-level belief that she has in mind, such as "that window will open" or "those people are being poisoned". The belief "I will intentionally open that window, unless I change my mind" can concern the same event as "that window will open", but its subject matter is altered, and qualified. Regardless, it is a mistake, in my view, to fixate too strongly on the conceptual connection between *belief* and *knowledge*, especially in relation to action[13]. Aims can have the very same objects as do ground-level beliefs, are warranted on the model of epistemic warrant, and have analogous "success conditions". They are the source of a kind of knowledge not just in accordance with the facts. Anscombe's conception of *practical knowledge*, at least, is therefore better accounted for within the boundaries of technical rationality.

13. Belief comes up only obliquely in *Intention*, and Anscombe seems at best doubtful about the prospects for reductive theories of phenomena such as intentional action or knowledge.

Once warranted aims are understood as (generally) embodying a kind of knowledge, one of the more subtle aspects of the relation between intention and belief falls into place. Some action theorists argue that the subject matter of an intention is necessarily general, while the subject matter of an apparently corresponding belief is *particular*. According to this view, Obadiah has a belief about a particular sandwich he witnessed being made, whereas Agatha's intention only concerns some general, potential sandwich. If this theory were right, an intention would be more like an existential generalization than a (ground-level) belief. However, although Anscombe's (perhaps facetious) metaphor of a "seeing *eye* in the middle of acting" is misleading, in that vision has the wrong direction of fit, the "grasping hand" of intentional action is nevertheless analogous to an eye in terms of how it can relate us to *future* particulars.

Davidson (1963, pp. 5–6) makes the argument about intention and particulars this way:

> We may be taken in by the verbal parallel between 'I turned on the light' and 'I wanted to turn on the light'. The first clearly refers to a particular event, so we conclude that the second has the same event as its object. Of course it is obvious that the event of my turning on the light can't be referred to in the same way by both sentences since the existence of the event is required by the truth of 'I turned on the light' but not by the truth of 'I wanted to turn on the light'. If the reference were the same in both cases, the second instance would entail the first; but in fact the sentences are logically independent. What is less obvious, at least until we attend to it, is that the event whose occurrence makes 'I turned on the light' true cannot be called the object, however intentional, of 'I wanted to turn on the light'. If I turned on the light, then I must have done it at a precise moment, in a particular way—every detail is fixed. But it makes no sense to demand that my want be directed to an action performed at any one moment or done in a unique manner. Any one of an indefinitely large number of actions would satisfy the want and can be considered equally eligible as its object.

Although Davidson writes here of *wanting* rather than *intending*, the passage occurs just before the introduction of his belief/desire pair theory, according to which the desire could be part of an intention. His observations therefore appear to apply just as well to a more neutral conception of turning on the light as an intended end, with flipping its switch as the intended means.

Either way, the "less obvious" part of this argument rests on a series of confusions. First, the belief "I turned on the light" can surely also have "any one of an indefinitely large number

of actions", intentional or not, as *its* object. To be something that one knows, that belief will typically relate to a particular illumination in virtue of its warrant—perhaps "I" witnessed the light turning on—although it may not[14]. Nothing in the content of either belief (or "sentence") alone picks out a particular turning-on. It can only do so, or fail to, with help. If the intention *cannot* do so, it must be because nothing can give an intention that kind of help.

One possible argument for there being no such help is that "every detail is fixed" in only the past and (perhaps) the present, while future details are changeable. I argue in Section 5.5 that the implications of this difference are less important than they might appear. Even setting that question aside, however, it is far from clear that the objects of true *beliefs* about the future are necessarily different than those about the past. Are the objects of "There will be an eclipse next Monday." said two weeks ago and of "There was an eclipse last Monday." said yesterday, necessarily different, just because the former is prospective? Does the fact that the eclipse may either be visible or be obscured by clouds weigh on this question? I don't see why they must be, or how it would.

It could also be argued that the will, free or not, precludes any relation to future particulars. Perhaps the movement of planets and moons is sufficiently "set" to connect us to certain future events, but if our actions were predictable in that way, they could not be the proper subject of an intention. The will, broadly speaking, clearly influences the relation of intention to its realization in at least one way: Because you can change your mind about even a high confidence intention, a given intention might forever lack any "object". To whatever extent our intentions can connect us to future particulars, their doing so is conditional on our following through on them. But this is not Davidson's worry. He argues that because we might act on an intention at one time in one way, or at another time in another way, there can be no referential connection between the intention and the event or circumstances that realize it.

The possibility of such a connection may be more familiar when considered in terms of the retrospective question, "What makes *this* illumination the intentional realization of *this*

14. Perhaps many lights are in another room, there are many switches here, only some of which work, and my only information comes from an illumination counter, which runs on a delay.

particular aim?". According to the theory of technical rationality, what makes the flipping intentional is the non-wayward realization of the aim to flip, which is the realization of *that* aim because of your awareness of it *as* you flip. And what makes the illumination intentional is the justificatory relation between the aim to flip and the aim to turn on the light, and the lack of any waywardness in *its* realization. If you had changed your mind about turning on the light, or momentarily forgot about that aim, but then acted on a whim to flip the switch, you would have flipped intentionally, but not in virtue of either of your original aims.

This is the same relation that Davidson appears to rule out, only traced from world to representation. An aim that is never intentionally realized, even if it is true, has no object in his sense[15]. But an aim that *is* intentionally realized has that particular realization as its object, however and whenever it happens. Just as a belief that depends on visual experience is connected to the object of that experience in virtue of the past operation of your eye, an aim that depends on a motor-control plan is connected to the object of that plan in virtue of the future operation of your hand. *That* peanut butter sandwich is therefore just as much the object of Agatha's aim as it is Obadiah's belief.

11.2.3 "The Cause of What It Understands"

I close with some thoughts on the relation between intentional action and the concept *cause*.

I have so far said little about Causalism. And although I lay out the inverse theory of Chapter 10 largely in terms of a taxonomy of "waywardness", its association with traditional Causalist theories is only circumstantial, because it does not depend on the existence of a "causal chain" between an aim and its realization. The theory may implicitly depend on certain causal chains, of course, but in this it is no different in that than almost every other theory that relates in any way to physical processes, including epistemology.

Must an intention be a cause of its realization, in order for the latter to be intentional? Is it not enough that the agent cause it? The thought that it must be a cause appears to come from

15. There might be many other events that make even an intentionally realized aim true. Agatha can intentionally make a peanut butter sandwich at the same table where a machine is simultaneously making dozens.

something like the following reasoning: If the intention has no influence whatsoever on the behavior that leads to its realization, then the behavior cannot be explained by that intention. It therefore must influence the behavior, which in physical terms means that it must be at least *a* cause of the corresponding bodily movements.

Even if this reasoning is correct, it poses the problem of how to interpret "in physical terms". The phrase "causal *chain*" suggests a sequence of events adjacent in time and space such that if one had a "God's eye view", one could trace the steps in a chain backwards and forwards in virtue of direct causal influence alone. This would be the kind of causation that might be defined in terms of physical properties, such as a transfer of energy or momentum. On this interpretation, Causalism would serve as a sort of physical check on theories of intentional action; if there is sometimes no causal chain between the supposed physical instantiation of an intention and the relevant behavior, the thought would go, that supposition is mistaken.

Given our current level of understanding of the neural basis of intentional action, it will be some time before this version of Causalism could serve as that check in practice. On the other hand, some of what we already know about physical computation counts against this theory. To be more efficient, quite ordinary CPUs sometimes *predict* the value of some input, and later discard any results of that prediction if it turns out to be wrong. Therefore, when the prediction is instead verified as correct, and certain other events are timed in certain ways, there will be no "physical causal chain" between the outputs based on the prediction and the verified value.

In the contemporary literature on causation, this would be called a "double-prevention"—the discarding of the results prevents their further use, and the verification of the prediction prevents the discarding (Hall 2004). This computational arrangement demonstrates that even all-or-nothing, deterministic *influence* is possible without a transfer of energy or momentum in some cases. And, tellingly, the standard illustrations of double-prevention and other varieties of *disconnection* use the metaphor of excitatory and inhibitory connections between neurons. Disconnection in neural processing is therefore not only possible, but a subtext of almost all recent discussion of the phenomenon.

Even setting the open question of the physical basis of intentional action aside, it is not difficult to construct an example of *subjective* double prevention, such as this variant of the excuse-offering pumpers of Anscombe (1957, §23–5):

> A man is employed to pump water into the cistern which supplies the drinking water of a house, and has done this almost every day for years. Someone has found a way of systematically contaminating the source with a deadly cumulative poison whose effects are unnoticeable until they can no longer be cured. This someone has calculated that if these people are destroyed some good men will get into power who will govern well, and has revealed the calculation, together with the fact about the poison, to the cistern-filler, who agrees.
>
> Every day after the man learns of the poison, the following happens: On his way to pump, he suddenly stops and remembers in shock that the water source has been poisoned. Then he remembers that poisoning the inhabitants is a good thing, and continues on to the pump, which he operates in awareness that the water is poisoned.

Because the man in this example pumps out of habit, there is little reason to think that the aim to poison is an "efficient" cause of his pumping in a narrow, physically defined sense. Instead, his awareness that the water is poisoned would ordinarily change his mind about pumping, but his awareness that the inhabitants should be poisoned counts against that change, and he does what he would normally do.

Even if the *influence* of an "input" on the "output" of a physical computation does not imply a transfer of energy or momentum, the Causalist intuition might still be vindicated on a more general interpretation of "in physical terms". The movement associated with the behavior is physical, and given some fairly ordinary assumptions, every intention has a corresponding, or identical, physical state. Causalism could therefore be the claim that the latter causes the former in a sense broader than the narrow physical definition, but not so broad as to fall under another of Aristotle's four categories. I take it that the philosophical debates summarized in Paul and Hall (2013) concern what should count as a cause in this sense. And it is plausible that physical computation or "information processing" implies that kind of causal relation.

Indeed, it *so* plausible as to call into question what the concept of *cause* can really tell us about the concept of *intentional action*. I agree with Schaffer (2000, p. 288) that each could

eventually be defined or grasped independently of the other. We might, for example, come to agree about what is and is not a cause in the context of physical computation without ascribing any intentional goals to the system. At our present degree of understanding, however, causal theorists benefit more from "Actionism" than action theorists benefit from Causalism. Rather than limiting what counts as a theory of intention, the intuition that an intention must be a cause of its realization is expanding our understanding of what counts as a cause.

If both this broader conception of Causalism and the inverse theory are correct, then the intentional realization of an aim or traim must be caused by it. That relationship is straightforwardly plausible in the case of a technical entitlement. Whether a basic aim or traim is realized intentionally depends on whether the neural motor control system and muscles function correctly. Even if that functioning does not imply physical causation, it amounts to a kind of information processing. And, as noted just above, the same explanation can account for a causal relation between a technically justified aim or traim and the intentional realization of its ultimately justifying basic aims or traims. The pumper above would not pump *this time* unless he aimed to poison. More generally, one acts on a justifying aim at least partly in virtue of having the aim it justifies.

Assuming a bit of transitivity, the remaining challenge is to explain how the realization of one's means must cause the realization of one's end—assuming these are numerically different—when the latter is intentional. Every theory of intentional action faces this problem, which cannot be reduced to physics alone. We might explain how the poisoning is a physical effect of the pumping in terms of the movement of poisoned water from the pump to the cistern, and then from the cistern into the bodies of the poisoned. But you can intentionally save my life by stepping in front of a bullet without there being any chain of physical causation between the stepping and myself, let alone "my life"[16]. If your movement "causes" my life to be saved, it is in some broader sense of that term.

16. This sort of "abstract causation" is one motivation for coarse-grained theories of action identity. If your blocking the bullet, or simply your stepping, *is* the saving of my life, then there is no further causation to explain. Taken to the extreme, this approach makes all intentional actions identical to certain mental events (Hornsby 1980, ch. III), which presumably leaves only neural causation to account for.

According to the inverse theory, an aim is realized intentionally when it is technically justified and realized without waywardness. But waywardness is described on the model of a Gettier problem, and therefore a lack of it does not imply the existence of a causal chain. Translated into the epistemic realm, the claim would be: If one knows that P and that Q, and Q is part of one's justification for P, and Q is a physical event[17], then Q either *causes* P or is *identical* to P. This is obviously false—Q could also be about some effect of P that counts as evidence, such as a shadow. So it may not be clear that the technical claim is even plausible—that it even has the right ingredients for truth.

What makes the inverse theory different is that technical reasoning sets up genuine counterfactuals. I argue in Section 7.3 that the common understanding of justification requires *relevance*. Relevance is neutral with respect to realm—if Q has no bearing on whether P, it cannot justify a belief that P—but it has less epistemic significance because one's inability to justify P without Q implies nothing about truth of Q. The likelihood that there will actually be a φing, in contrast, very much depends on whether one aims to φ. When you reason that the truth of A depends on φing, and that φing depends on *you*, you are not mistaking φ for some effect of A, or some effect of an event that also causes A. If it were those, then φ would be true regardless of your other aims and beliefs.

Technical reasoning is not certain, of course. Some other person or thing might φ instead, or A might come about some other way. Almost all counterfactual judgments are probabilistic. Alternative ways that one's end might be realized would therefore pose a problem … if they were not exactly what the requirement against waywardness rules out. The theory does not need to explain a causal influence of intentionally realized means on *any* realization of the end, only an *intentional* realization. The inverse theory therefore supports this claim about every non-basic intentional action: An agent's technical reasoning establishes a counterfactual relationship between means and end, and subsequent events "vindicate" that reasoning, in the way that a lack of a Gettier problem vindicates an epistemic justification, making what is justified count as knowledge.

17. Generalizing a bit from the class of things that can count as a basic action.

I will not argue that this claim implies a causal (or identity) relation, only that there are three plausible standards on which it might. The most liberal would be in virtue of the counterfactual relation alone. The most conservative would add the representation of the counterfactual relation, and the requirement that that representation cause the realization of the means. By the counterfactual standard, a doubly-prevented change in the course of a rock in a landslide would be a cause of that course. It would count by the representational standard only if rock 3 were thrown by a person or system that "understood" how it hitting rock 2 would prevent that from hitting rock 1. An intermediate standard would appeal to teleology. If a system does not represent a counterfactual, but its good functioning still somehow depends on it, we might attribute a causal relation when we would not otherwise. As usual, function could either be accounted for in terms of explicit design, as with tools, or evolutionary history, as with organisms.

The intentional realization of an aim is its effect by any of these standards, but the narrowest has special significance for another of Anscombe's claims about practical knowledge. The parallel structure of epistemic and technical rationality, and their analogous success conditions, makes it appropriate to think of aim as the basis of practical knowledge: As with a belief, an aim is an attitude of distinctly high confidence, and can be held appropriately only when it is warranted. By justifying an end aim with aims or traims, one takes the latter as means to that end, and thereby *understands*, in light of other justifying beliefs, how the realization of the means will bring about the realization of the end. When, and only when, the end is realized without waywardness, we can think of the aim as embodying practical knowledge, as justified true belief without a Gettier problem embodies epistemic knowledge.

So suppose it does turn out to be right that, in virtue of the "glue" of justification and the logic of technical counterfactuals, an aim is also a *cause* of its intentional realization. In that case, we could rightly follow Aquinas and Anscombe (Anscombe 1957, §48) by observing that an aim is "the cause of what it understands".

APPENDIX A

A Summary of the Theory

As *belief* is the attitude towards a proposition that it is true, *aim* is the attitude towards a proposition that one will realize it, absent a change of mind, and therefore that it is *voluntarily true*. Each is an attitude of high confidence, with the high *credence* of belief corresponding to the high *technical confidence* of aim. Either type of confidence is appropriate only if it is *warranted*. There are two types of warrant: *entitlement* and *justification*.

Epistemic entitlement to a *perceptual belief* is in virtue of the reliability or proper functioning of the perceptual system that produces the perceptual experience from which it is derived. *Technical entitlement* to a *basic aim*—an aim to perform a manual task t under a certain condition w—is in virtue of a prospective assessment of high technical confidence—grounded in the relevant capacity for movement—that one can perform the action assuming that w holds.

A belief that P can be *epistemically justified* by other warranted beliefs [or suspicions] in virtue of patterns of reasoning that "preserve" truth or high likelihood. An aim that E can be *technically justified* by at least one technically warranted aim [or traim] and any number of epistemically warranted beliefs [or suspicions] in virtue of the same patterns of reasoning. Technical justification is the instantiation of the means-end relation; in technically justifying an aim that E in part by an aim that M, one takes M as a means to E.

The epistemic attitude of *suspicion* associates a degree of credence c_e with some proposition P. A suspicion that P should either be justified by warranted beliefs or suspicions in virtue of (often probabilistic) reasoning that supports P to degree c_e, or carry a lower-credence entitlement to P in virtue of some perceptual capacity (as when seeing something at a distance). The technical attitude of *traim* associates a degree of technical confidence c_t with some propo-

sition *E*. A traim that *E* should be justified by at least one warranted aim or traim and any number of warranted beliefs or suspicions by the same kinds of (often probabilistic) reasoning that support *E* to degree c_t. Or a traim to *t* when *w* should carry a lower-technical-confidence entitlement to *t* when *w* in virtue of the prospective assessment of confidence provided by the relevant capacity for movement (as when trying to hit a distant target). Attempts to realize an aim that *E* will typically be described by the agent as "doing", while attempts to realize a traim that *E* will typically be described as "trying".

A belief that *P* (or not-*P*) or suspicion (c_e,*P*) must generally be *voluntarily recalled* when the question "whether *P*?" becomes relevant. A basic aim (*w*,*t*) or basic traim (c_t,(*w*,*t*)) must generally be *non-voluntarily recalled* when one has epistemic warrant for the belief that *w* holds. A non-basic aim that *E* must generally be non-voluntarily recalled when the high technical confidence in it becomes unwarranted, to allow for re-planning or changing one's mind about pursuing *E*.

Just as an epistemically warranted belief must be true to count as *knowledge*, a technically warranted aim must be true to count as an *intentional action*. But truth is not sufficient in either case.

A true perceptual belief with epistemic entitlement can fail to be knowledge if it is derived from a veridical hallucination or not properly derived from any perceptual experience. A true basic aim with technical entitlement can fail to be an intentional action if the capacity for the movements that make it true is not functioning properly, or the movements do not result from an exercise of will in awareness of the aim.

An epistemically justified true belief can fail to be knowledge if it suffers from a *Gettier problem*. A technically justified true aim can fail to be an intentional action if it suffers from *consequential* or *tertiary waywardness*. Each of these types of waywardness is the analogue of a Gettier problem, with the consequential variety tracing to a justifying aim and the tertiary variety tracing to a justifying belief. An aim that comes true without the involvement of the agent will also fail to be intentional on her part, given a Gettier-style analysis. Suspicions and traims can suffer related problems, but while a true suspicion is not considered to be

knowledge, a technically warranted true traim is an intentional action as long as it does not suffer from waywardness.

BIBLIOGRAPHY

Adams, Frederick. 1986. "Feedback About Feedback: Reply to Ehring." *The Southern Journal of Philosophy* 24 (1): 123–131.

———. 1995. "Trying: You've Got to Believe." *Journal of Philosophical Research* 20:549–561.

Anscombe, G. E. M. 1957. *Intention.* 2nd ed. Harvard University Press, 2000.

———. 1979. "Under a Description." *Noûs:* 219–233.

Audi, Robert. 1973. "Intending." *The Journal of Philosophy* 70 (13): 387–403.

Binkley, Robert. 1965. "A Theory of Practical Reason." *The Philosophical Review:* 423–448.

Block, Ned. 1995. "On a Confusion About a Function of Consciousness." *Behavioral and Brain Sciences* 18 (2): 227–247.

Bratman, Michael. 1984. "Two Faces of Intention." *The Philosophical Review* 93 (3): 375–405.

———. 1999. *Intention, Plans, and Practical Reason.* The David Hume Series. Cambridge University Press.

———. 2009. "Intention, Belief, and Instrumental Rationality." In *Reasons for Action,* edited by David Sobel and Steven Wall, 13–36. Cambridge University Press.

Brown, Harriet, Rick A. Adams, Isabel Parees, Mark Edwards, and Karl Friston. 2013. "Active Inference, Sensory Attenuation and Illusions." *Cognitive Processing* 14:411–427.

Brunero, John. 2010. "Self-Governance, Means-Ends Coherence, and Unalterable Ends." *Ethics* 120 (3): 579–591.

Burge, Tyler. 1993. "Content Preservation." *The Philosophical Review* 102 (4): 457–488.

———. 2003. "Perceptual Entitlement." *Philosophy and Phenomenological Research* 67 (3): 503–48.

———. 2005. "Disjunctivism and Perceptual Psychology." *Philosophical Topics* 33 (1): 1–78.

———. 2013. *Cognition Through Understanding: Self-knowledge, Interlocution, Reasoning, Reflection.* Vol. 3. Oxford University Press.

Cath, Yuri. 2015. "Revisionary Intellectualism and Gettier." *Philosophical Studies* 172 (1): 7–27.

Chisholm, Roderick. 1966. "Freedom and Action." In *Freedom and Determinism,* edited by Keith Lehrer. Random House.

Davidson, Donald. 1963. “Actions, Reasons, and Causes.” Reprinted in Davidson 2001, 3–19, *The Journal of Philosophy* 60 (23): 685–700.

———. 1971. “Agency.” In *Agent, Action, and Reason,* edited by Ausonio Marras, Richard Bronaugh, and Robert Binkley, 1–37. Reprinted in Davidson 2001, 43–61. University of Toronto Press.

———. 1973. “Freedom to Act.” In *Essays on Freedom of Action.* Reprinted in Davidson 2001, 63–82. Routledge.

———. 1978. “Intending.” In *Philosophy of History and Action.* Reprinted in Davidson 2001, 83–102. D. Reidel Pub. Co.

———. 2001. *Essays on Actions and Events.* Philosophical Essays Series. Clarendon Press.

Doody, Rachelle Smith, and Joseph Jankovic. 1992. “The Alien Hand and Related Signs.” *Journal of Neurology, Neurosurgery & Psychiatry* 55 (9): 806–810.

Enç, Berent. 2004. “Causal Theories of Intentional Behavior and Wayward Causal Chains.” *Behavior and Philosophy:* 149–166.

Flegal, Kristin E., and Michael C. Anderson. 2008. “Overthinking Skilled Motor Performance: Or Why Those Who Teach Can’t Do.” *Psychonomic Bulletin & Review* 15 (5): 927–932.

Fleming, Stephen M., Raymond J. Dolan, and Christopher D. Frith. 2012. “Metacognition: Computation, Biology and Function.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367 (1594): 1280–1286.

Frankfurt, Harry G. 1978. “The Problem of Action.” *American Philosophical Quarterly* 15 (2): 157–162.

Fried, Itzhak, Amiram Katz, Gregory McCarthy, Kimberlee J. Sass, Peter Williamson, Susan S. Spencer, and Dennis D. Spencer. 1991. “Functional Organization of Human Supplementary Motor Cortex Studied by Electrical Stimulation.” *The Journal of Neuroscience* 11 (11): 3656–3666.

Gallagher, Shaun. 2006. *How the Body Shapes the Mind.* Clarendon Press.

Gettier, Edmund. 1963. “Is Justified True Belief Knowledge?” *Analysis* 23 (6): 121–123.

Gibbons, John. 2001. “Knowledge in Action.” *Philosophy and Phenomenological Research* 62 (3): 579–600.

Gollwitzer, Peter M. 1993. “Goal Achievement: The Role of Intentions.” *European Review of Social Psychology* 4 (1): 141–185.

Gollwitzer, Peter M., and Paschal Sheeran. 2006. "Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes." *Advances in Experimental Social Psychology* 38:69–119.

Hall, Ned. 2004. "Two Concepts of Causation." In *Causation and Counterfactuals,* edited by John Collins, Ned Hall, and Laurie Paul, 225–276. MIT Press.

Harman, Gilbert. 1976. "Practical Reasoning." *The Review of Metaphysics:* 431–463.

Holton, Richard. 2008. "Partial Belief, Partial Intention." *Mind* 117 (465): 27–58.

Hornsby, Jennifer. 1980. *Actions.* International Library of Philosophy, Psychology, and Scientific Method. Routledge & Kegan Paul.

———. 2008. "A Disjunctive Conception of Acting for Reasons." In *Disjunctivism: Perception, Action, Knowledge,* edited by Adrian Haddock and Fiona Macpherson, 244–261. Oxford University Press.

Horsey, Richard. 2002. "The Art of Chicken Sexing." *UCL Working Papers in Linguistics* 14.

Jeannerod, Marc. 2001. "Neural Simulation of Action: A Unifying Mechanism for Motor Cognition." *Neuroimage* 14 (1): S103–S109.

Knobe, Joshua. 2003. "Intentional Action and Side Effects in Ordinary Language." *Analysis* 63 (279): 190–194.

Kolodny, Niko. 2007. "State or Process Requirements?" *Mind* 116 (462): 371–385.

Korsgaard, Christine M. 2008. "Acting for a Reason." In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology,* 207–29. Oxford University Press.

———. 2009. *Self-constitution: Agency, Identity, and Integrity.* Oxford University Press.

Lehrer, Keith, and Thomas Paxson. 1969. "Knowledge: Undefeated Justified True Belief." *The Journal of Philosophy* 66 (8): 225–237.

Levy, Yair. 2013. "Intentional Action First." *Australasian Journal of Philosophy* 91 (4): 705–718.

Lewis, David. 1980. "Veridical Hallucination and Prosthetic Vision." *Australasian Journal of Philosophy* 58 (3): 239–249.

Luthra, Yannig. 2013. "Embodied Rational Agency." PhD diss., University of California, Los Angeles.

———. 2016. "Non-rational Aspects of Skilled Agency." *Philosophical Studies* 173 (8): 2267–2289.

MacKenzie, Donald. 2008. *An Engine, Not a Camera: How Financial Models Shape Markets.* MIT Press.

Marušić, Berislav. 2013. "Promising against the Evidence." *Ethics* 123 (2): 292–317.

McBride, Jennifer, Petroc Sumner, Stephen R. Jackson, Nin Bajaj, and Masud Husain. 2013. "Exaggerated Object Affordance and Absent Automatic Inhibition in Alien Hand Syndrome." *Cortex* 49 (8): 2040–2054.

Mele, Alfred R. 1987. "Intentional Action and Wayward Causal Chains: The Problem of Tertiary Waywardness." *Philosophical Studies* 51 (1): 55–60.

———. 2003. "Intentional Action: Controversies, Data, and Core Hypotheses." *Philosophical Psychology* 16 (2): 325–340.

Mele, Alfred R., and Paul K. Moser. 1994. "Intentional Action." *Noûs* 28 (1): 39–68.

Millgram, Elijah. 2016. "Practical Reason and the Structure of Actions." In *The Stanford Encyclopedia of Philosophy,* Summer 2016, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2016/entries/practical-reason-action/.

Morrison, John. 2016. "Perceptual Confidence." *Analytic Philosophy* 57 (1): 15–48.

Moss, Sarah. 2013. "Epistemology Formalized." *Philosophical Review* 122 (1): 1–43.

Nelson, Thomas O., and Louis Narens. 1990. "Metamemory: A Theoretical Framework and New Findings." *Psychology of Learning and Motivation* 26:125–173.

O'Grady, Jane. 2001. "Elizabeth Anscombe." Published January 11, 2001, *The Guardian.* https://www.theguardian.com/news/2001/jan/11/guardianobituaries.highereducation.

O'Shaughnessy, Brian. 1973. "Trying (As the Mental 'Pineal Gland')." *The Journal of Philosophy* 70 (13): 365–386.

Pacherie, Elisabeth. 2008. "The Phenomenology of Action: A Conceptual Framework." *Cognition* 107 (1): 179–217.

Parsons, Terence. 1989. "The Progressive in English: Events, States and Processes." *Linguistics and Philosophy* 12 (2): 213–241.

Paul, L. A., and Ned Hall. 2013. *Causation: A User's Guide.* Oxford University Press UK.

Paul, Sarah K. 2009a. "How We Know What We're Doing." *Philosophers' Imprint* 9 (11): 1–24.

———. 2009b. "Intention, Belief, and Wishful Thinking: Setiya on "Practical Knowledge"." *Ethics* 119 (3): 546–557.

Paul, Sarah K. 2010. "Deviant Formal Causation." *Journal of Ethics & Social Philosophy* 5:1–23. http://www.jesp.org/articles/view.php?id=56.

Raz, Joseph. 2005a. "Instrumental Rationality: A Reprise." In *Journal of Ethics & Social Philosophy | Symposium I,* 1–19. http://www.jesp.org/articles/downloadSymposium.php?id=13&comment=5.

———. 2005b. "The Myth of Instrumental Rationality." *Journal of Ethics & Social Philosophy* 1 (1): 28.

Ross, Jacob. 2009. "How to Be a Cognitivist About Practical Reason." *Oxford Studies in Metaethics* 4:243–281.

Roth, Abraham Sesshu. 2000. "The Self-referentiality of Intentions." *Philosophical Studies* 97 (1): 11–51.

Scanlon, T. M. 2007. "Structural Irrationality." Chap. 4 in *Common Minds: Themes From the Philosophy of Philip Pettit,* edited by Geoffrey Brennan, Robert Goodin, Frank Jackson, and Michael Smith. Clarendon Press.

Schaffer, Jonathan. 2000. "Causation by Disconnection." *Philosophy of Science* 67 (2): 285–300.

Schroeder, Mark. 2009. "Means-end Coherence, Stringency, and Subjective Reasons." *Philosophical Studies* 143 (2): 223–248.

Schroeder, Severin. 2001. "The Concept of Trying." *Philosophical Investigations* 24 (3): 213–227.

Schwenkler, John. 2015. "Understanding 'Practical Knowledge'." *Philosophers' Imprint* 15 (15).

Searle, John R. 1975. "A Taxonomy of Illocutionary Acts." In *Language, Mind and Knowledge,* edited by K. Gunderson, 344–369. Reprinted in Searle 1985, 1–29. Minneapolis: University of Minnesota Press.

———. 1983. *Intentionality: An Essay in the Philosophy of Mind.* Cambridge University Press.

———. 1985. *Expression and Meaning: Studies in the Theory of Speech Acts.* Cambridge University Press.

Setiya, Kieran. 2007a. "Cognitivism about Instrumental Reason." *Ethics* 117 (4): 649–673.

———. 2007b. *Reasons Without Rationalism.* Princeton University Press.

———. 2008. "Practical Knowledge." *Ethics* 118 (3): 388–409.

———. 2009. "Practical Knowledge Revisited." *Ethics* 120 (1): 128–137.

———. 2012. "Knowing How." *Proceedings of the Aristotelian Society* 112 (3): 285–307.

Shepherd, Joshua. 2017. “Halfhearted Action and Control.” *Ergo* 4 (19): 259–276.

Sinhababu, Neil. 2013. “The Desire-Belief Account of Intention Explains Everything.” *Noûs* 47 (4): 680–696.

Snedegar, Justin. 2017. *Contrastive Reasons.* Oxford University Press.

Thomson, Judith Jarvis. 1971. “The Time of a Killing.” *The Journal of Philosophy* 68 (5): 115–132.

Turri, John. 2016. “Knowledge Judgments in ‘Gettier’ Cases.” In *A Companion to Experimental Philosophy,* 335–348. Wiley.

Valaris, Markos. 2014. “Instrumental Rationality.” *European Journal of Philosophy* 22 (3): 443–462.

van de Snepscheut, Jan L. A. 1993. *What Computing Is All About.* Springer-Verlag New York, Inc.

Velleman, David. 1992. “What Happens When Someone Acts?” *Mind* 101 (403): 461–81.

———. 2007. *Practical Reflection.* David Hume Series. CSLI Publications.

Vogler, Candace. 2009. *Reasonably Vicious.* Harvard University Press.

Way, Jonathan. 2012. “Instrumental Rationality.” In *Routledge Encyclopedia of Philosophy,* edited by Tim Crane and Roger Crisp. Taylor & Francis. https://www.rep.routledge.com/articles/thematic/rationality-instrumental/v-1.

Williamson, Timothy. 2017. “Acting on Knowledge.” In *Knowledge First: Approaches in Epistemology and Mind,* edited by J. Adam Carter, Emma C. Gordon, and Benjamin Jarvis, 163–181. (text page references are to pre-print version). Oxford University Press.

Wittgenstein, Ludwig. 2010. *Philosophical Investigations.* John Wiley & Sons.