UNIVERSITY OF CALIFORNIA SAN DIEGO

# A Defense of Hybrid Voluntarism

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

in

Philosophy

by

Cory Michael Davia

Committee in charge:

Professor David O. Brink, Chair
Professor Richard J. Arneson
Professor Craig R.M. McKenzie
Professor Dana Kay Nelkin
Professor Manuel Rogelio Vargas

2020

The Dissertation of Cory Michael Davia is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

TABLE OF CONTENTS

VITA

| | |
|---|---|
| 2010 | Bachelor of Arts, Claremont McKenna College |
| 2012 | Master of Arts, University of Wisconsin, Milwaukee |
| 2020 | Doctor of Philosophy, University of California San Diego |

ABSTRACT OF THE DISSERTATION

A Defense of Hybrid Voluntarism

by

Cory Michael Davia

Doctor of Philosophy in Philosophy

University of California San Diego, 2020

Professor David Brink, Chair

Some of our talk and thought is descriptive; it attempts to say how the world is. Some of our talk and thought is normative; it attempts to say how the world should be. This dissertation addresses the relationship between those two domains. Specifically, I investigate the question: in virtue of what do some descriptive considerations have the normative status of reasons for action? Philosophers working on this question have tended to defend three possible answers: that considerations are reasons in virtue of mind-independent normative facts, in virtue of the desires of the agents for whom they

are reasons, or in virtue of the voluntary choices or acts of will of the agents for whom they are reasons. I argue that the best answer is a hybrid of the first and third answers. While it is not always up to us whether some consideration is a reason, sometimes it is. Investigating how this can be turns out to illuminate some classic metaethical issues about the role of reflection in agency, the relationship between morality and our personal projects, and how normative motivation is possible.

# Chapter 1

## Creating Reasons

Everyone has a reason not to harm others, but only some people have a reason to make a careful study of ethical theory. Everyone has a reason to appreciate art, but only some people have a reason to learn blues harmonica. Everyone has a reason to eat their vegetables, but only some people have a reason to try out the trendy new restaurant with a vegetarian tasting menu.

The above examples show that reasons can have different scopes; some apply to everyone and some only apply to some of us. This much is uncontroversial. More controversial, though, are claims about what explains why there are these differences in our reasons. One explanation is that what we have reason to do is sometimes up to us. When we decide to engage in particular projects or be particular kinds of people (say, a moral philosopher, a band member, or a cook), we acquire reasons we wouldn't otherwise have. Over the next several chapters, I will argue that we should take this thought at face value. Sometimes, by act of will, we can create reasons.

In this chapter, I'll make this claim more precise and start laying the foundation for an argument for it. I'll first explain what kind of reasons I have in mind, then what it would be to create them in the sense I have in mind. The view I'll end up defending is one according to which what we have reason to do is sometimes – but only sometimes – up to us. It's sometimes up to us because some reasons metaphysically depend on our wills, but it's only sometimes up to us because there are also reasons that do not.

## 1. Disambiguating Reasons

The term "reasons" is notoriously ambiguous, so some taxonomy is in order here. We can start by distinguishing *practical reasons* from *theoretical reasons*, where the former are reasons to do things and the latter are reasons to believe things. When I say that I think that it's sometimes up to us what reasons we have, I mean practical reasons.[1] So, the kinds of cases that will concern us going forward are ones where agents are deciding what to do.[2]

Practical reasons might be further divided into three categories. Suppose that I decide to go to Twiggs coffee shop. One thing we might want to know is what counted in favor of doing that. This is a question about *justifying reasons*. Any number of things might contribute to explaining why that was a worthwhile thing to do: they serve delicious biscuits there, getting out of the house sometimes helps me focus on writing, it'll give me a chance to pick up groceries on the way home. Another thing we might ask is what explains why I made the decision I did. This is a question about *motivating reasons*.[3] It could be that the same things that make it reasonable to go to Twiggs also explain why I did it, but these can come apart. Maybe what motivated me to

1 This is not to deny that sometimes what we ought to believe is up to us as well, but I won't take a stand on that here. Chakravartty (2017) develops a view according to which some questions about what to believe, e.g. about whether to believe in the unobservables postulated by a scientific theory rather than just in the observable predictions the theory makes – depend on some voluntary choices. Callahan (ms) defends a more general form of epistemic voluntarism.

2 There is a debate about how to characterize what practical reasons count in favor of, strictly speaking. One option is that they count in favor of doing things, another option is that they count in favor of bringing about states of affairs. My sympathies are with the former and I'll assume it here, but I think what I say could be translated without loss if necessary. See Hurley (2018) for discussion about the implications this choice has for ethical theory.

3 These kinds of reasons are often thought to be what's at the heart of the difference between action and mere bodily movement. Actions, the thought goes, are things for which there are motivating reasons. Though, there's disagreement about what kind of explanation is at issue here. The standard view, usually attributed to Davidson (1963), is that motivating reasons are species of causal explanations. Other philosophers have thought that these are teleological explanations, and at least some of those have conceived of those explanations in a non-causal way. See Sehon (2016) for discussion. I won't take a stand on that disagreement here.

go was the thought that Twiggs is a funny name, even though that doesn't really count in favor of going there (I can appreciate the sound of it from home, after all).[4] Finally, we might want to know what, if asked, I would cite as a justification for this decision. This is a question about what *reasons I have in mind*. Here again, this might line up with what actually justifies going or what actually explains my going, but it might not. Perhaps I think that I made this decision in order to get some writing done, but really what drove my decision was the hope that I'd run into friends there.[5]

My primary concern here will be justifying reasons. I'll be defending the claim that we can make it the case that something counts in favor of an action when it wouldn't have otherwise. Nonetheless, these other senses of "reason" will still be relevant. Because these senses can overlap, each sets some constraints on how we theorize about the others. Justifying reasons should at least typically be the sort of thing that could motivate us and we could have in mind while acting, even if in day to day life these often come apart.[6]

A few more distinctions within justifying reasons will be helpful. It often happens that while there's something to be said for an option, it's nonetheless not what one should do all things considered. I'll mark this difference by calling a reason *decisive* if it settles what to do, and

4 Motivating reasons explain actions by being the sorts of things that guide our actions, or for the sake of which we act. But this isn't the only kind of explanation we might be interested in. If I misread the social situation and make a bad joke, you might explain my doing so by pointing out that I was anxious, even though anxiety isn't what motivated me to make the joke.

5 There are, I think, analogues of these distinctions for theoretical reasons: considerations that justify a belief might be different from the ones an agent would cite when asked why she believes, which might also be different from the psychological story about why she believes. Still, I'll just be talking about practical reasons here.

6 The "at least typically" qualifier is important. For one thing, there's a debate about whether justifying reasons will motivate any agent who recognizes them. See for instance Brink (1997a) and Smith (1994). I lean towards thinking they needn't, but I won't weigh in on that here. For another, there are complications about whether agents with the types of psychologies we have can always act in light of the relevant reasons directly. See note 13 below.

*pro tanto* if it's just one of many reasons that need to be assessed together in order to reach a verdict.[7] For example, it might be that the taste of the biscuits at Twiggs is a *pro tanto* reason for me to order one, but the fact that I've already had breakfast is a decisive reason not to. My claim about creating reasons is intended in the *pro tanto* sense; whether the reasons we create settle what we ought to do depends on what other reasons there are.

Justifying reasons can also be sorted by the people to whom they apply. When it's my mother's birthday, that's a reason for me to call her, but not for you to do so. Other reasons have a wider scope. When the World Series is starting, that's a reason for baseball fans to tune in, but not for non-fans. Still other reasons, like our reasons to eat our vegetables or avoid harming others, have a maximally wide scope. We can call reasons with less-than-maximal scope *agent-relative*, and reasons everyone shares *agent-neutral*.[8] On the view I'll defend, when we create

7 There are further distinctions one might make about how reasons interact in order to reach a verdict about what to do. For instance, one *pro tanto* reason might preempt another instead of outweighing it, by making it inappropriate to consider that reason in these circumstances. That it's your birthday might normally count in favor of giving you the biggest slice of cake, but if the biggest slice of the cake contains the only antidote to the poison we've all been exposed to, then it would be inappropriate to take into account whose birthday it is when deciding who gets which piece. It could also be that reasons count in favor of options in different ways. Dancy (2004) suggests that while some reasons count in favor of actions by making it rational to do them, others make options more appealing without thereby making one rationally criticizable for not doing them. For my purposes here, we can get by with the distinction between *pro tanto* and decisive reasons, but questions like these about how reasons combine to produce verdicts about what to do will be relevant in Chapters 3 and 6.

8 I've drawn the distinction between agent-relative and agent-neutral reasons in terms of their scope. Other philosophers have drawn it differently. Nagel (1986) p152-153, for example, draws the contrast in terms of the "general form" or principle behind particular reasons. For instance, suppose that I call my mother because it will make her happy. Plausibly, this is an application of the principle that when we can make others happy, we should. That principle doesn't make any essential reference to me; if you could make my mother happy by calling her, you'd have a reason to do that too. We might contrast this with a case where I call Valerie Davia because she's my mother. Plausibly, this is an application of the principle that we each ought to call our mothers. That principle does make essential reference to individuals; I'm supposed to call *my* mother, not yours.

In many cases, my way of drawing this distinction and Nagel's way will overlap. When there are reasons I have that you don't, they'll correspond to principles that make reference to features I have and you don't. Nonetheless, the overlap might not be perfect. Consider a principle that says everyone ought to be generous. This principle seems not to make any essential reference to features of particular agents, but if what I am in a position to give is different from what you are in a position to give, then applying this principle will result in my having reasons that you don't. So, these will turn out to be agent-relative reasons on the way I'm drawing the distinction and

reasons they'll be reasons specifically for us.

Finally, when talking about the reasons that bear on a decision, we sometimes take in to account the epistemic situation of the person deciding. Suppose that, unbeknownst to me, Twiggs is out of biscuits. In cases like this, we might say that *subjectively* – that is, given what I take the facts to be – the availability of biscuits is a reason for me to go, while *objectively* – that is, given the actual facts of the situation – it's not. Similarly, we might sometimes want to relativize our talk about reasons to the normative perspective of the person acting. If someone (falsely) thinks that it's wrong to sell biscuits, then we might say that subjectively the presence of biscuits on Twiggs' menu is a reason for her not to patronize Twiggs, while objectively it's not.[9]

We can now put my claim more precisely (and more awkwardly): I think that some of our

---

agent-neutral reasons on the way Nagel does.

I prefer the way of articulating this distinction in the main text because I am not sure how to decide what the general form of a particular reason is. Suppose I give some friends some home-brewed beer because I have some around. Is the general form of this reason a generic principle about being generous, or an agent-referencing principle about sharing whatever it is you're in a position to share? My difficulty with Nagel exegesis aside, I don't think this difference will be important for my project here. For some purposes, e.g. trying to convince moral skeptics that they're implicitly committed to recognizing some agent-neutral reasons, it might be really important which reasons are agent-relative and which are agent-neutral. For my purposes, I don't think the differences will be important. It's just important to have a way of flagging that some considerations count in favor of my doing things, without counting in favor of your doing those things (or even helping me do those things).

9 What I'm calling "objective" reasons are what Parfit (2011) calls "fact-relative" reasons. They're the reasons there are, however things may appear from my epistemic perspective. What I've called "subjective" reasons are what Parfit calls "belief-relative" reasons. Parfit also discusses "evidence-relative" reasons, i.e. reasons there appear to be given my evidence. These will differ from my belief-relative reasons whenever my beliefs are not appropriately responsive to my evidence. One could go further and introduce other kinds of relativizations, e.g. reasons I have in light of the evidence that I should have had, or could have discovered. I'm happy to admit that these are all perfectly good senses of "reason," and that for certain philosophical purposes they're useful, especially when assessing agent's actions under imperfect epistemic conditions. I stick to "objective" and "subjective" in the main text merely because I don't think the arguments I'll go on to make depend on these differences. Some philosophers also distinguish between the objective reasons there are and the subset of those that an agent has epistemic access to by calling the latter reasons she "has." See Schroeder (2008). I won't be using the language of possession in this specialized way; when I say an agent has a reason, I'll just mean that there is a reason relevant to her decision. Instead, I'll describe situations where an agent does not have epistemic access to all of her objective reasons as cases where her objective and subjective reasons come apart.

objective practical justifying agent-relative reasons are self-created, and that these make a *pro tanto* difference to what we ought to do. I'll do my best to flag these distinctions when they're relevant, but when I use "reason" in an unqualified way, this is the sense I mean.

These distinctions might make our ordinary use of the term "reason" seem like a bit of a mess. But, I think there is more unity here than meets the eye. As Mark Schroeder points out, many of these different kinds of reasons can be analyzed in terms of the objective, *pro tanto*, agent-relative kind. Agent-neutral reasons can be thought of as a special case of agent-relative reasons, where the conditions of application are very inclusive (for instance, they might be reasons that apply to us insofar as we are agents, or people, or members of the moral community).[10] Subjective reasons can be thought of as purported reasons, the kinds of things that look like objective reasons within a given set of epistemic constraints.[11] Similarly, one can construct an account of decisive reasons by starting with *pro tanto* reasons and providing a procedure for comparing and aggregating them.[12]

Similar points apply to the relations between justifying reasons, motivating reasons, and the reasons we have in mind when we act. When an agent is fully informed and not self-deceived, the reasons she will cite just are her motivating reasons. And when she's ideally rational, she's motivated by all and only things that justify.[13] This isn't to say that real life agents

10 Alternatively, one might think of agent-relative reasons as agent-neutral reasons that apply in very specific circumstances. But see Schroeder (2007a) for an argument that the arrow of explanation goes the other way.

11 Schroeder (2007b) p10-21

12 Articulating such a procedure turns out to be very tricky. Issues about how to compare and aggregate reasons will be important in Chapter 3 and Chapter 6.

13 This claim might need some qualification. Markovits (2014) p38-49 suggests a few cases where doing what you ought to do requires not attending to the reasons that justify it. For example, suppose that you are on plane and the pilot is incapacitated. You can land the plane, and the danger that your fellow passengers will be in if you don't is a very strong reason for you to do so. Nonetheless, attending to this reason while trying to land the plane might unnerve you. So, perhaps it's better for your motivating reasons and justifying reasons to come apart here, even if you're ideally rational.

like us meet these conditions – we're motivated by all kinds of stuff, and we've never fully informed about the world or ourselves – but still we can see these other kinds of reasons as interesting applications of the justifying kind.

So, it looks like the concept of an agent-relative, objective, *pro tanto* reason is central to our ordinary talk about reasons. And talk about reasons is plausibly central to much of our talk about other normative concepts like well-being and value. To say that something is good for me or valuable is – at least in part – to say that I have reason to care about or pursue it.[14] It would be illuminating, then, to have a theory about how these reasons do their justificatory work.

## 2. The Metaphysics of Reasons

Suppose that, while I'm thinking about whether to go Twiggs, I find out that a friend is already working there. Suppose further that this is one of my reasons (in the justifying sense) for deciding to go. One thing we might ask is, why is this a reason for me?

This question admits of at least two readings. On one reading, it's a request for further reasons, ones that justify thinking that the presence of my friend really justifies going to Twiggs. Someone asking this question might agree with the descriptive claim that my friend is at Twiggs, but nonetheless wonder whether that really counts in favor of going. So, we might answer this question by pointing out that it's fun to spend time with friends, or that we can keep each other accountable while we work. In other words, we can answer this question by pointing to further reasons.

14 See Brink (2008) p7-8. Brink notes, and I agree, that one can recognize these connections without committing one way or the other about which notion is more fundamental; perhaps value should be explained in terms of reasons, perhaps it's the other way around, or perhaps neither has explanatory priority. See the Postscript to this chapter for some further discussion.

On another reading, this question is a request for metaphysical explanation. Someone asking this question might agree with the normative claim that my friend's presence is a reason for me to go to Twiggs, but nonetheless wonder in virtue of what that's true. So, to answer this question we'd have to identify the metaphysical grounds in virtue of which my friend's presence is a reason for me to go.

The difference between these two questions can be understood by analogy with questions about causation in philosophy of science. It's one thing to want to know what causes what, and another to want to know in virtue of what one things causes another. So, knowing the full story about what causes what still might not tell us everything we want to know about causation. Similarly, having a fully worked out theory about what reasons there are might not tell us everything we want to know about reasons for action.

This second question has gone by a handful of names. Christine Korsgaard calls it a question about the source of a reason's normativity,[15] or what one needs to add to a non-normative description in order for it to get its normative status as a reason (analogously, one might wonder what, in addition to correlation, is required for causation). Schroeder frames efforts to answer this question as attempts to see how reasons fit in with a naturalistic picture of the world.[16] Julia Markovits characterizes the first question as one about "what reasons we have" and the second as one about "what reasons are."[17]

These characterizations all seem to be in the same ballpark, but in what follows I am

15 Korsgaard (1996). Korsgaard initially sets this up as a question about explaining the normativity of moral reasons specifically, but in subsequent work – see especially Korsgaard (2009) – her focus shifts to practical reasons more generally.
16 Schroeder (2007b)
17 Markovits (2014)

going to stick with calling this a question about the metaphysics of reasons, or a question about in virtue of what some consideration is a reason. I'm hesitant about the "source of normativity" language both because of exegetical questions about Korsgaard – it's sometimes hard to tell when she's addressing our first question and when our second – and because philosophers have used the word "normativity" in a variety of ways that make it easy for us to talk past each other.[18] I'm hesitant about Markovits' framing because, as Schroeder points out,[19] it might turn out that what explains why something is a reason is not itself part of that reason, in the same way that inauguration ceremonies are part of what makes someone a president, but are not themselves part of any president.[20] Asking what reasons are makes it very tempting to look only for something that our reasons might be identical to, but I don't want to prejudge what kinds of answers we can give by assuming that the metaphysical ground of some reason will also be identical to that reason. Finally, I'm hesitant about Schroeder's framing because, as I'll discuss below, some philosophers think that reasons can't be made to fit in with the rest of our picture of the world, and yet they seem to be addressing the same question Schroeder is.

So, putting things in terms of metaphysical explanation seems like the most neutral option. Going forward, it'll be easier to formulate different views about this if we stipulate one last bit of terminology. Take the sorts of descriptive claims that might or might not function as reasons, things like "My friend is working at Twiggs," "Twiggs is fun to say," or "there are biscuits at Twiggs." One could take different views about what these potential reasons are;

18 While Korsgaard clearly has justificatory force or authority in mind, in other contexts "normativity" is sometimes thought of as a motivating force, or something more descriptive like a statistical average.

19 Schroeder (2007b) p24

20 Another way to put this point is that the bearers of normativity (the things that have justifying force) might not be the same as the source of normativity (the thing in virtue of which they have that justifying force). See Chang (2009) p244.

maybe they're facts, sentences, propositions, states-of-affairs, beliefs, or what have you.[21] To stay neutral on this question, I'll use the intentionally vague term *considerations*.[22] This allows us to put our question by asking, in virtue of what does a given consideration count as a reason?

## 3. Theories about the Metaphysics of Reasons

So, what does it take for some consideration to be a reason? Here are some possible views.

One possible answer is that nothing is required. T.M. Scanlon and Derek Parfit, for example, defend a view we might call *reasons primitivism*, where "reason" is an unanalyzable concept, not explicable in non-normative terms and the thing in terms of which all other normative talk is to be explained.[23] On this view, some considerations just are reasons, and some just aren't. We might be able to explain the justifying force of some reasons in terms of other reasons (like, when we explain my reason to go to Twiggs in terms of a reason to spend time with my friends), but there's no explanation available that doesn't ultimately bottom out in reasons.

A similar answer is that a consideration gets to be a reason with the help of some other normative concept, like prudential, aesthetic, or moral value. While "there are biscuits at Twiggs" just describes the world, it might come to justify going to Twiggs in conjunction with the idea that a good life involves gustatory pleasure, or that the biscuits are beautifully crafted, or that it would be kind to buy one for the next person in line. On this view, we *can* explain why

21 This is one place where it matters that justifying reasons and motivating reasons should be able to overlap; whatever we choose here should be the kind of thing that can plausibly play both roles. See note 3 above and Alvarez (2010) for discussion.

22 I'm borrowing this way of fudging things from Scanlon (1998).

23 See Scanlon (2013) and Parfit (2011).

some consideration is a reason, but the explanation stays within the broader arena of normative discourse.[24] On both of these types of views, then, there can be answers to the normative reading of the question "why is this consideration a reason?" but no answers to the metaphysical reading of it.

I'm going to group these two views together under the name metanormative *externalism* – "externalism" because reasons are not to be explained in terms of some feature of the agents who have them (or at least, not only in terms of such features), and "metanormative" to distinguish this externalism from the myriad of others in philosophy.[25]

A more radically different answer is that we can explain what makes a consideration a reason in terms of the desires, pro-attitudes, or affective states of the agent for whom that consideration is a reason. Perhaps what makes "my friend is working at Twiggs" a reason for me to go there is that I have a desire to spend time with my friends, and this consideration identifies a way in which I can promote the satisfaction of this desire.[26] I'll call this type of view metanormative *internalism*.

What makes these views so different is that according to internalists, we can explain something normative (reasons) in terms of some non-normative things (considerations and

24 Some recent advocates of this view include Dancy (2000), Shafer-Landau (2003), Wallace (2006), Wedgwood (2007), Brink (2008), and Enoch (2011a).

25 Enoch remarks that internalism and externalism are "the most widely abused terms in philosophy." Enoch (2011a) p302. I'm inclined to agree.

26 Some recent advocates of this view include Brandt (1979), Williams (1979), Velleman (2000), Railton (1986), Schroeder (2007b), and Markovits (2014). Such views are often presented as a way of elaborating on ideas in Hume (1738).

desires).[27] So, internalism is a reductive theory in a way that externalism is not.[28] But despite this difference, internalism and externalism share an important feature. Because our desires are not under our voluntary control (we might over time inculcate some and try to rid ourselves of others, but at any given moment we just have the desires we have), both views have it that our reasons are not up to us.

A third view, we'll be calling it *voluntarism*, denies this. Voluntarists agree with internalists that we can explain why a consideration is a reason by appealing to some feature of the agent for whom it's a reason, but for voluntarists the relevant feature is an agent's will. Since what we will is up to us, it's a consequence of this view that we can make something a reason just by deciding to, much like how (when our bodies cooperate) we can move around physical space just by deciding to.[29]

One way to get your head around this view is to compare it to divine command theory. On some conceptions of God, God has the power to make a consideration a reason by holy decree. So, for example, it might be that what explains why "lobster is a shellfish" is a reason not to eat lobster is that that's what God decided.[30] We can think of voluntarism as a secular analogue

27 This contrast might sound familiar. In debates about moral reasons specifically, the view I'm calling "externalism" sometimes goes by the name "realism." If you'd like, feel free to make that substitution. I'm hesitant to make it, though, because while internalists explain the reasons we have in terms of some feature of the agent, it needn't follow from this that our reasons are dependent on us in the way realists characteristically deny. For example, Schroeder (2007b) develops a sophisticated version of internalism where our reasons are fundamentally explained by our desires, but we can nonetheless have reasons to do things we don't want to do, and fail to have reasons for what we do want to do. He takes his view to be the best version of realism. So, rather than join the fray over who counts as a realist, I'm going to avoid that label here. The interesting disagreement here doesn't seem (to me, at least) to be about whether reasons are real, but about what makes them real.

28 If you have reasons primitivist sympathies, you might take the reductive ambitions of internalism to suggest that these views are really just talking past each other. Perhaps internalists are just stipulating a definition of reason, and the real question is which sense of reason is the one we care about when talking about justifying action. I think this is too hasty, but I'll return to this question in Chapter 3.

29 It might sound non-obvious that what we will is up to us. That's an artifact, I think, of the term "will" being used for lots of different purposes in philosophy. I'll say more below about the particular sense I have in mind.

30 See Brink (2006) for critical discussion of this view.

of this view.[31] It says that each of us has this power to decide which considerations are reasons for us. We do so by willing that those considerations be reasons.[32] For example, I might make the fact that there's a lacuna in the literature a reason for me to write a paper by committing to being an academic. If I'm not an academic, then that kind of consideration is just neither here nor there for me.[33]

## 4. Voluntarism: Subtleties and Clarifications

The idea of an agent's "will" has been used for various purposes in the history of philosophy, so to give a more precise gloss on voluntarism I need to say more about how I intend to be using that term here. For my purposes, what's importantly distinctive about the voluntarist thesis is that it (a) identifies something non-normative in virtue of which considerations count as reasons, and (b) that thing is voluntary for the agents whose reasons those are. This puts some constraints on how we can understand the nature of the will. For example, (a) rules out an account according to which the will is a faculty by which we respond to the normative facts. If that's what the will is, voluntarism turns out to be a version of externalism. Similarly, (b) rules out an account according to which the will is a particular type of desire, like a desire about which

31 Note that voluntarism is the secular analogue of the view that it's God's *will* that can make some consideration a reason. Other views that put God at the center of the metaphysics of normativity might go differently. For example, it might be that what makes "lobster are shellfish" a reason not to eat lobster is that God wants us to refrain from eating shellfish. Because that view ultimately grounds normativity in a desire rather than an act of will, it's an analogue of internalism rather than voluntarism.

32 Korsgaard (2009) is the foremost contemporary defender of this view. Historical antecedents for it can be found in Hobbes (1651) and Pufendorf (1672). Korsgaard claims Kant for this tradition as well, but that's controversial; I'm going to stay out of questions about Kant scholarship here.

33 The possibility of voluntarist views might might make "internalism" sound like a poor choice of name for views that explain reasons in terms of desires. After all, wills are just as internal to agents as desires are. I agree, and if we were starting the debate between these views from scratch I might choose a different name. But, there's a long history of discussion of internalism of this sort, so for the sake of continuity with it I'm going to stick with that name.

other desires to have. If that's what the will is, voluntarism turns out to be a version of internalism.[34] Of course, this needn't mean that we don't have such mental states or capacities, they're just not the ones I mean to be talking about when I talk about the will.[35]

As far as I can see, this leaves us with two options. One option is to understand the will as a capacity for reflective identification, i.e. as an ability to step back from the first-order facts about you and to decide which of them you identify with. On this view, willing something is a matter of having a particular higher-order attitude, one of identification with that thing. This sounds high-brow, but I think it's a pretty familiar way of thinking about at least some kinds of deliberation. Ruth Chang helpfully illustrates with the example of a stock Hollywood romantic comedy.[36] Jack likes his uncommitted bachelor life, but falls for Jill. We, the audience, are frustrated that Jack won't "commit." What do we want him to do? Plausibly, at least one of the things we want is for him to change his reflective stance toward his feelings for Jill. We want him to see those feelings as part of who he is rather than as obstacles to living his uncommitted bachelor life. We may also want him to do other things, like promise to care for Jill or to work to change some of his habits, but these things typically follow – and would strike us as insincere

34 Though it's worth noting that this would be a type of internalism that's pretty different from contemporary versions of that view (see Chapter 3 for a tour of the possibilities). So, even if you accept this account of what the will is, the arguments that follow should still be interesting to you. They'll just be arguments for a distinctive kind of internalism. Thanks to Amy Kind for pressing me on this.

35 In this way, I'm departing from some other philosophers' usage. For instance, Herman (2007) thinks of the will as a capacity to be motivated by duty, and Frankfurt (1971) uses "will" to pick out the desire of an agent that succeeds in leading that agent to action. Of course, nothing I've said amounts to a reason for thinking that philosophers like these are wrong to use the term "will" in the way they do. If you'd prefer to reserve the term "will" for something else, you can plug another term into the voluntarist thesis. I just need a word for the conjunction of (a) and (b) above and using "will" is continuous with the literature. For what it's worth, I think all of these ways of using the term "will" are more about theory-guided stipulation than they are about natural language. For instance, to say that someone is "strong-willed" isn't to say anything about their capacity to choose, to be guided by duty, or the structure of their second-order desires.

36 Chang (2013a) p5-6

without – this change in Jack's will.[37]

Note that while this conception of the will appeals to the concept of "identification," it's not the kind of identification that Harry Frankfurt famously spells out in terms of higher-order desires.[38] When we want Jack to commit, we don't just want his desires to care for Jill to be motivationally efficacious. We want, instead, something about Jack's stance toward those desires to change. In particular, we want him to see them as relatively central to who he is. For example, if Jack wanted his desire to care for Jill to lead him to action, but he also saw those actions as relatively disconnected from his plans for his life or the kind of person he is (i.e. if he regarded them the way he regards his desire to watch TV to pass the time), this wouldn't strike us as a movie-script ending. Compare: I like the Star Wars movies, I want to see new ones when they come out, and I want that desire to lead me to action, but that's not sufficient for being a committed Star Wars fan. Being a fan involves a richer connection between those desires and the rest of what I like and do, e.g. going to premieres in costume, having a theory about whether Admiral Holdo's suicide mission opens up a plot hole in the overall series narrative, collecting memorabilia, debating the most narrative-satisfying order to watch the movies in, and so on. In short, liking Star Wars is not as central to my web of commitment as it is for Star Wars fan. Similarly, for Jack to commit to Jill in the way we root for in romantic comedies, his feelings for Jill need to occupy a privileged place in his mental economy.[39]

37 In this case, Jack commits to a person: he wills that facts about Jill's projects and well-being be reasons for him. Other paradigm instances of this kind of commitment concern careers, hobbies, and other sorts of personal projects.

38 Frankfurt (1971)

39 Of course, integration with the rest of one's pursuits isn't all or nothing. One might be less or more of a Star Wars fan. My claim here is just that for Jack to commit in the relevant sense, Jill can't be on the periphery.

A second option is to treat the will as a sui generis ability of agents.[40] Specifically: the ability to choose or to form intentions, where choosing isn't just a matter of being moved by some desire or other. On this view, the will can't be defined in terms of some other mental activity (like reflectively endorsing or identifying); instead, the ability to choose is just one of the fundamental kinds in an agent's mental life, like belief or desire. If so, acts of will work much like other basic actions. Just as I can raise my arm just by deciding to do it, without going through any intermediary actions, it might be that willing is just one of the things I can do without going through any intermediary mental steps. What's going on in the Jack and Jill case, on this picture, is that there's some basic mental action we want Jack to do, one that rationalizes and typically gives rise to the other kinds of changes we expect from him, e.g. changes in his habits and his self-conception.

I'm partial to the first of these options, so going forward I'm going to describe willing in much the same way I described Jack's commitment. But, I don't think the arguments I'll go on to make depend on this choice. Instead, what will matter is just the more minimal claim that the will is some psychological feature of agents that is under their voluntary control, however exactly we want to pick out what that psychological feature is. So, voluntarism is the thesis that voluntary choices of this kind are what explain why considerations are reasons.

Two further clarifications are in order about the sort of voluntariness I have in mind. First, I don't mean to imply that changes in agents' wills are always consciously brought about. Jack might drift along liking his life with Jill, and one day find himself having changed his

40 Chang (2009) seems to have the former conception of the will in mind; in conversation she has expressed some sympathy for the latter.

reflective stance toward that liking. I don't, however, take this undermine the thought that the change was voluntary. Compare: I might find myself tapping my foot without noticing it, but this doesn't change whether foot-tapping is under my voluntary control. Second, while changes in our wills are up to us, the various phenomena that tend to go along with them might not be. For instance, I might commit to an exercise regime, willing that "it's Tuesday" count in favor of going to the gym, but fail to muster the motivation to act on those new reasons, fail to have them be salient in deliberation, and so on. After this happens a few times, I might give up my commitment. So, there is some sense in which it's not up to us whether all of our commitments stick; the world and the rest of our psychology have to cooperate. But this, I think, doesn't threaten the idea that whether we can make the commitments in the first place is up to us, and it's that voluntary act that voluntarists say metaphysically explains reasons.

To sum up: I'm understanding the will as a particular mental state agents can have that is under their voluntary control. This voluntary control distinguishes willings from normative beliefs (what I believe is constrained by my assessment of the evidence) and from my desires (while I might exercise some long-range control over my desires, at any given time I just have the ones I have). I'm hoping to stay neutral on exactly what kind of attitude fills this conceptual space – maybe it's a sui generis one or maybe it's a matter of reflective identification – but I also hope that the kind is a recognizable one, and that by gesturing at cases I've at least provided an intuitive push toward thinking that we do have such mental states. In any event, voluntarist views about reasons explain our reasons in terms of this kind of thing.

To help draw out what's unique about this view, it'll help to contrast the kind of control

voluntarists think we have over our reasons from the kind of control recognized by other views. Suppose that I have a desire to keep my promises, or that one of the brute facts about reasons is that I have a reason to keep my promises. If so, I can end up with a reason to pick you up at the airport by promising to pick you up at the airport. This is a special case of a more general point: what we do has downstream effects on what's the case, and what's the case is relevant to what reasons we have. Nonetheless, according to both internalism and externalism, what explains why I have a reason here is not my act of promising all by itself; it's my act of promising plus the relevant desire or normative fact. In contrast, voluntarism says that whether some consideration is a reason (not just whether some circumstance obtains) is what's up to us. According to voluntarism, we can explain my reason to pick you up solely in terms of my choice, rather than in terms of my choice and some other feature of the situation.[41]

This observation helps us categorize some otherwise similar-sounding views. Suppose, for example, that one of the agent-independent normative facts is that exercises of will are valuable. On a view like this, I can change the reasons I have by exercising my will. Is this view therefore voluntarist? At least on the way I'm carving things up here, the answer is "no." According to voluntarism, the will is the metaphysical ground of reasons. But according to this other view, what grounds my reasons is my will *plus* the agent-independent fact about the value

41 This contrast parallels one in the literature on "normative powers." Watson (2009) characterizes them broadly as abilities to change the norms or entitlements that one (in the case of a promise) or another (in the case of an order, gift, or permission) is subject to. On this characterization, the ability to change the non-normative facts might count as a normative power, since some non-normative changes will give rise to normative ones. My promising to pick you up at the airport might change your expectations in a way that would make it wrong for me to violate them. Alternatively, following Owens (2012), one might think of normative powers more narrowly as abilities to change norms or entitlements directly. My promise to pick you up at the airport might transfer authority over whether I go to the airport from me to you.

of acts of will. So, I count this view as externalist.[42] A similar point applies to the relationship between externalism and internalism. Suppose that the only agent-independent normative fact is that fulfilling desires is valuable. On a view like this, my reasons will be a function of my desires. Is this view therefore internalist? Again, the answer for my purposes is "no" – internalism is a thesis about the metaphysics of reasons rather than just about what reasons there are.

## 5. Looking Ahead and Methodological Assumptions

Philosophers who defend each of these views about the metaphysics of reasons typically take them to be views about all reasons, but nothing about the question we're asking precludes a hybrid answer; there might be more than one way for a consideration to come to be a reason. For example, Chang defends a view according to which some of our reasons are given to us in the way externalists think that all reasons are, but that these reasons don't always conclusively settle what we ought to do. When they don't, Chang thinks, we can create reasons in the way voluntarists think all reasons are created.[43]

I agree with Chang that the best account of the metaphysics of reasons is a hybrid between externalism and voluntarism, and my main goal in what follows is explain why this is plausible.[44] Nonetheless, I'm skeptical about whether Chang's account is the most plausible

42 See Chapter 6 for more discussion of this kind of externalist view.

43 Chang (2013b). The details of Chang's view are a bit more complex than this; I'll discuss them in more detail in Chapter 5.

44 One might read Nagel (1986) as agreeing, at least in broad outline, with Chang about this too, though I am unsure if that's the right reading. On the one hand, Nagel does seem to write in a way that's friendly to a hybrid voluntarist reading, e.g. when he acknowledges "unassailable pockets of subjectivity" (p155) within an objective framework for thinking about value, and when he discusses "reasons of autonomy" that we have in virtue of our choices and commitments, and which force a distinction between "what can be valued from an objective standpoint and what can be seen from an objective standpoint to have value from a less objective standpoint" (p166). On the other hand, one might read these as normative claims about what has value or what people have

version of hybrid voluntarism. In particular, I think there is room for hybrid voluntarists to disagree about the ways our voluntarist reasons (considerations that are reasons in virtue of our willing) are conditioned and constrained by our externalist reasons (considerations that are reasons in virtue of mind-independent normative facts). So, in addition to my arguments for hybrid voluntarism as a family of views, I'll also set out some alternative versions and suggest some reasons one might be attracted to them. So, my overall goal here is to move us from a debate about *whether* we can create reasons to a debate about *when* we can do so.

Here's a sketch of how things will go. In the next two chapters, I'll raise some objections to purebred voluntarist views (Chapter 2) and purebred internalist views (Chapter 3). I'll consider a range of objections to each view, but most prominently I'll argue that neither of these views can accommodate our ordinary judgments about what reasons people have. In Chapter 4, I'll offer a qualified defense of externalism, arguing that the usual objections to recognizing agent-independent normative facts don't go through. This tells us two things: that hybrid voluntarism need not be objectionable for having an externalist component, and that our choice of views can be narrowed to purebred externalism and hybrid voluntarism. In Chapter 5, I'll summarize some of the arguments for hybrid voluntarism already in the literature and offer some novel defenses of their more controversial premises. This discussion will provide a jumping off point for getting more specific about the details of my preferred hybrid voluntarist view, and how it compares to the one Chang defends. In Chapter 6, I'll put forward some new arguments for preferring hybrid voluntarist views. In particular, I'll explain how hybrid voluntarism provides the resources to

reason to do, rather than metaphysical claims about in virtue of what a given consideration is a reason. Nagel's earlier endorsement of realism (p138) suggests the latter reading. I'm of two minds about how to resolve this tension, but if it turns out that Nagel is my ally here, I am happy to have him.

give us better explanations of normative motivation, the range of ways agents can rationally disagree with each other about their reasons, and our relationships to our personal projects. Finally, in Chapter 7, I'll step back and identify some reasons one might be skeptical about any form of hybrid voluntarism, and try to assuage those worries.

At a macroscopic level, this plan has two main argumentative moves. First, I'll argue that purebred versions of internalism and voluntarism don't account for our intuitive judgments about the reasons we have. Second, I'll argue that, among the views that can account for them, hybrid voluntarism does so in the most satisfying way. Each of these main moves requires a background assumption about theory choice, so let me now put my methodological cards on the table. Here are two issues and two promissory notes.

First, in arguing against pure forms of internalism and voluntarism, I'm going to be assuming that a metaphysical theory about what it is in virtue of which some considerations are reasons should not force us to be very revisionary in our normative theory about what reasons we have. This assumption mirrors one we typically make about normative theorizing: a theory about ultimate moral principles (like the categorical imperative) should not force us to be very revisionary in our judgments about particular cases (like whether to lie to the inquiring murderer). This much I take to be uncontroversial; we have to start theorizing somewhere, and our intuitions about cases are part of what we have to go on. More controversial, though, are judgments about how revisionary it's okay for a theory to force us to be. After all, our intuitions about cases may be inconsistent, incomplete, or biased, so it should be open to defenders of general theories (be they normative or metaphysical theories) to argue that some of our intuitions

should be reformed rather than accommodated.[45] Moreover, theory choice is contrastive, so losses when it comes to accommodation might be justified by gains in some other dimension. I don't have much to say about the general methodological question about how to balance accommodation and reform in our intuitions about cases. As it turns out, though, I think the metaphysical aims of the disputants in the debate I'm engaging with here put extra pressure on us to be wary of reformist strategies. Here, then, is the first promissory note. When I'm making arguments that hinge on a theory's ability to accommodate intuitions about cases – especially in Chapter 3 when I'm arguing against internalism – I'll be putting them in the context of those extra pressures. In doing so, I think I can be fairly ecumenical about the more general story about the relationship between principles and intuitions in ethics.

Second, in arguing for hybrid voluntarism over purebred versions of externalism, I'm going to be assuming that assessing a theory's ability to accommodate intuitions about cases is not all there is to assessing that theory. As we've seen already, two candidate theories about the metaphysics of reasons might overlap in their implications about cases. For instance, internalism overlaps with an externalist theory according to which the only mind-independently good things are states of affairs in which our desires are fulfilled. Moreover, theories might do worse at accommodating intuitions precisely because what motivates them is some other concern, e.g. about ontological parsimony. So, to choose among theories we're going to need to appeal to desiderata that do not have to do with what the theory implies about particular cases. For the most part, I think what I say about this will be non-controversial. I'll rely on familiar desiderata

45 See Brink (2014) for some discussion of how this plays out in the context of the history of ethics and contemporary worries about what social science tells us about our normative intuitions.

(e.g. ontological parsimony, unity, fecundity, and explanatory power) and on the idea that a theory of the metaphysics of reasons should fit well with our other philosophical theories that implicate reasons (e.g. our theories about what it takes to act for a reason and how agents are motivated). Nonetheless, in Chapter 6 I'll rely on some more controversial assumptions about the relationship between metaphysical and normative theorizing. It'll be easier to spell out the desiderata in the context of the debate between hybrid voluntarists and purebred externalists, so here's the second promissory note. I'll flag and motivate my (potentially idiosyncratic) metaphilosophical assumptions in Chapters 6 and 7.

## 6. Postscript: Skepticism about Reasons

In this chapter, I have tried to set the stage for a discussion of the metaphysics of reasons. I'm certainly not alone in framing questions about normative discourse this way. Over the last fifteen years or so, it's become increasingly popular to frame traditionally metaethical questions in terms of reasons, rather than in terms of morality specifically. One advantage of this is that it helps make clear how these questions are often just as apt when asked about non-moral norms. For instance, just as a moral skeptic might ask what reason she has to be moral, a prudence skeptic might ask what reason she has to take her future interests into account. And just as we might ask what justifies our beliefs about morality, we might ask what justifies our beliefs about aesthetics or the standards of rationality.[46] Another advantage of focusing on reasons is that it helps to highlight the relations between questions in metaethics about how actions are justified

46 Of course, that we can ask the same questions about non-moral norms doesn't imply that any given answer will be equally plausible for any given set of norms. Questions about the authority of rationality, for instance, might be answered in ways that are not available for questions about the authority of morality. I'll return to this issue in Chapter 2.

and questions in action theory about how agents do things for reasons. So, I think, there is something illuminating about the generality of "reasons" talk.

Still, this way of talking about metaethics has its detractors.[47] So, while in the chapters that follow I'll mostly be engaging with other philosophers who set things up in similar ways, I'll try to say something here about why I don't think reasons talk is likely to lead us astray.

Some philosophers who put metaethical questions in terms of reasons do so because they think that the concept of a reason is the fundamental normative concept, the one in terms of which all other normative concepts (like goodness, obligation, etc) should be understood.[48] If that's right, then an account of the metaphysics of reasons could be extended to be an account of the metaphysics of the normative more generally (and similarly for the epistemology of reasons, or any other metanormative question you might want to ask). I think this claim is true, but it isn't obviously true; it could be that some other normative concept is fundamental, or that normative concepts hang together and no single one is most basic.

So while I accept it, strictly speaking I hope to avoid committing to this fundamentality claim here. Consider the examples with which this chapter started, about the reasons one might acquire by becoming a moral philosopher, blues band member, or cook. I said that a natural way of thinking about those examples is that they illustrate a contrast between times when an agent's reasons are not up to her, and times when they are. But this thought could be put other ways: we might say that the constituents of a person's *good* are sometimes up to her, or that whether

47 See Wedgwood (2015), Fogal (2016), Broome (2018) and Titelbaum (2017) for some arguments against the "Reasons First" program, some of which I will engage with below. A related challenge concerns whether there's *any* normative concept that's more fundamental than the others; see Wodak (ms) for skepticism about whether we should put any normative concept first.

48 See, for example, Scanlon (2013) and Parfit (2011).

something is *valuable* is sometimes up to her, or that the features of the world that explain what she *ought to do* are sometimes up to her. I could then go on to claim, as I did about reasons, that this observation is not only true but metaphysically deep. Sometimes we create our good, or value, or whatever. The question I'm most interested in here is about the relationship between the normative and the voluntary, so while I find framing this question in terms of reasons helpful, I'd be happy to give up that framing if it were shown that looking at reasons isn't the right way to get a grip on the normative.

I suspect, though, that this disavowal will not be very satisfying to philosophers who are skeptical about reasons. They might think that even if, in principle, I could restate my claims in terms of some other normative concept, thinking about things in terms of reasons will lead us to take certain intuitions too seriously or to confuse importantly different phenomena. So, it's worth looking at a couple of arguments for skepticism about reasons fundamentality and seeing whether they pose a threat to the way I've set things up here. I won't try to engage every possible argument against reasons fundamentality, but I hope that what I do say provides some grounds for optimism that my using reasons talk won't lead us too far astray.[49]

One argument against reasons fundamentality comes from Ralph Wedgwood. He notes that the idea of a justifying reason is asked to play two different roles in normative theorizing: reasons are both the kinds of things that explain what agents ought to do, and they are the kinds of things to which agents respond if they are ideally rational. Wedgwood thinks that no single thing could play both of these roles.[50] If this is right, then it looks like my project here is doomed

49 See Schroeder (2018) for some further discussion.
50 Wedgwood (2015)

to fail; we won't be able to answer questions about the metaphysics of reasons if nothing actually plays the reasons role.

Wedgwood motivates this skepticism by appeal to some examples. One has to do with action in cases of uncertainty.[51] Suppose that a friend and I are about to order brunch, but she has to leave the table to answer a phone call. Suppose also that we can't linger over brunch, so I ought to order when the waiter first comes around. We've eaten at this restaurant before, and we've talked about our tastes in brunch food, so I'm pretty confident that my friend will want the lemon ricotta pancakes. Still, I'm not certain about it; maybe she's decided to try something new, or maybe she had pancakes yesterday and is tired of them. In a case like this, Wedgwood says, what explains why I ought to order the pancakes is the fact that my friend wants the pancakes, while what – if I'm rational – I'll respond to is not this fact, but instead some other fact about the probability that she does, or some cognitive state of mine like my hunch that she does. Either way, it looks like it's not one and the same thing that explains why ordering pancakes is correct and why ordering pancakes is what a rational agent would do.

Wedgwood sees similar problems in cases where the facts that explain why you ought to do something are complicated.[52] Suppose that you kill someone in self defense. It's likely that the things you're responding to when you make that decision – even if you're fully rational – are not the things that a completely worked out moral theory will say actually explain the permissibility of that killing. For instance, you might have thought that killing was permissible because your attacker was culpable for her attack and because you posed no risk to bystanders, when in fact it

51 Wedgwood (2015) p9-11
52 Wedgwood (2015) p11-14

was permissible because of those things *and* your attacker's relatively high chances of succeeding in killing you. In these kinds of cases too, it looks like there can be distance between what explains why you ought to do something and what a rational agent would take into account. Even rational agents can't take *everything* into account.

These examples raise good questions about the relationship between justifying reasons and motivating reasons, but I am not sure that I see deep problems here. In the uncertainty case, it seems open to a fan of reasons to say that one thing really does play both roles – the fact that my friend wants pancakes – it's just that appropriately responding to this fact is not necessarily a matter of forming an outright belief about it. Similarly in the self-defense case, it seems open to fans of reasons to say that what ultimately explains the permissibility of your action isn't one really complicated reason that you can't grasp, but the interaction of several more easily graspable ones. If so, we can still say that the considerations you responded to are genuinely part of the explanation of why what you did was permissible. But I don't want too much to hang on the details of these cases; perhaps Wedgwood could design others that are harder to finesse. More importantly, the difficulties here are ones that come up when we think about the epistemic limits of the agents in the examples; we don't always know all the facts or which of the facts are relevant. For my purposes here, I think we'll be able to safely focus on cases where it's clear what to do and why, then ask in virtue of what those things are true. So, even if focusing on reasons has the potential to mislead in cases like Wedgwood's, I don't think it'll mislead us here.

Another challenge comes from Daniel Fogal. His worry starts with the observation that "reason" functions as both a count noun ("there are three reasons to go to Twiggs") and a mass

noun ("you have more reason to go to Twiggs than to stay home"). He observes that in other cases where a word can be used in both of these ways, the mass noun tends to be the more fundamental. For example, if you want to tell someone what the count noun "lights" means, you say that they are the kinds of things that emit the mass noun "light" – it'd get things backwards to say that light is the kind of thing that comes out of lights. Fogal takes this as evidence for thinking that "reasons" in the *pro tanto* sense I've focused on are not fundamental. Instead, the notion of what you ought to do is fundamental, and considerations get to be reasons by being cited in relation to that.[53]

Fogal notes that if this diagnosis of the relationship between the count noun "reasons" and the mass noun "reason" is correct, then projects like mine that rely on intuitions about when it's felicitous to say that some consideration is a reason for some agent are going to face some messy data. If (count) reasons are just what's cited in explanations of all things considered judgments about what we ought to do, then we would expect our intuitions about when they're appropriate to depend on conversational context. For example, suppose that there will be a crab cake festival in Baltimore and that Michelle loves crab cakes When explaining why Michelle ought to go to Baltimore, I might say "she loves crab cakes" or "there will be crab cakes there." If I say just one of these things, either of them sounds like a reason for Michelle to go. But if I say both, you'll rightly suspect that something fishy is going on. This suggests that our judgments about what's a *pro tanto* reason are unstable, sensitive to what else has been said rather than offering insight into the underlying metaphysics. In particular, we might hesitate to draw conclusions about when agents' psychological states are or are the ground of reasons for them,

53 Fogal (2016) p5-7

since intuitions about these things seem to depend on conversational context.[54]

Fogal finds the linguistic data more impressive than I do. For one thing, it's not clear to me that when words can be used as both a count noun and a mass noun, we should default to thinking that the mass noun is more fundamental. Light seems that way, but other cases not so much. For instance, Fogal says it's clear that we would not explain what pleasure is in terms of individual pleasures; instead, pleasures are the kinds of things that give us pleasure.[55] That seems backwards to me. The mass noun sense of pleasure is very disunified. Playing pool, drinking beer, and going to philosophy talks all give me pleasure, but I'd be hard pressed to identify any overlap in their phenomenology. I'm tempted, then, to run the explanation the other way, and say that there are lots of individual things we recognize as pleasures, and that pleasure is just an abstract way of talking about the myriad feelings we have when we do those individual things. So, absent further argument, I think we should take the relationship between the count and mass senses of "reason" as up for grabs as well.[56]

Moreover, while I agree with Fogal that we should be careful about conversational context when pumping intuitions about reasons, I don't think the situation is as dire as he makes it out to be. Notice that we don't react to examples like Michelle's by saying "Huh, this

54 Fogal notes that these could be different reasons in the right circumstances, like if in addition to loving crab cakes Michelle gets paid a hefty sum for each crab cake she eats. But, in ordinary circumstances, these two facts each sound like reasons on their own but together sound like one reason. Fogal (2016) p15-18

55 Fogal (2016) p5-7

56 This commits me to rejecting the account of pleasure in Bentham (1789), where there are only quantitative differences between pleasures, not qualitative. The cases above seem like clear counterexamples to me. Note, though, that one could go along with my claim about the qualitative disunity of pleasure without thereby accepting the view that some kinds of pleasures are strictly more valuable than others, as Mill (1863) does. This might be true too, but it's a further claim; some pleasures could be different from others without thereby being better. Even if we focus on a fixed quality of pleasure, there can be differences. For example, going to philosophy talks and going to art museums are both plausibly higher pleasures but don't feel the same, while eating pizza and watching low-brow TV are both plausibly lower pleasures but don't feel the same.

consideration seemed like a reason and now it doesn't – what gives?" Instead, we think "Huh, this consideration seemed like a reason, but it seems like the *same* reason as that other consideration – what gives?" The first reaction would leave us in the dark about how to proceed from intuitions about cases, but the latter opens the door to some more careful theorizing. We might, for example, think that it's the conjunction of those considerations that's a reason for Michelle, and that they seemed like reasons individually only because each was understood to be elliptical for the conjunction. Or maybe that Michelle loves crab cakes is part of the background conditions that explain why the presence of crab cakes on a particular occasion is a reason for her.[57] In short, rather than leaving us with nothing to go on, Fogal's observations about the instability of reasons-judgments give us more data we can use to test theories about the metaphysics of reasons. A good theory should explain both why these things sound like reasons individually, and why when they're together there's a suspicion of double-counting rather than some other kind of infelicity like a violation of Gricean norms.

Of course, what I've said in response to Wedgwood and Fogal isn't decisive, and there are other worries about reasons fundamentalism in the literature.[58] But, I hope I have said enough to suggest that framing metanormative questions in terms of reasons is not a lost cause, and that attention to the worries of reasons skeptics can help inform the best way to do it. This, at the very least, bolsters my confidence that even if ultimately it turns out that reasons talk is not the best way to frame metanormative questions, I can help myself to it here without a terribly distorting effect. In any case, an argument needs to start somewhere, and in what follows I'll start with

57 Schroeder (2007b) explores this possibility in Chapter 2. Chang (2009) p257 suggests a similar idea: individuating reasons both by their content (the consideration that serves as a reason) and their source (whatever gives that consideration its normative force).

58 See, for example, Titelbaum (2017) and Broome (2018).

intuitions about reasons.

# Chapter 2

## Problems for Purebred Voluntarism

My goal in the chapters that follow is to defend hybrid voluntarism, i.e. the idea that some – but only some – considerations are reasons in virtue of our wills. As a result, I have two kinds of interlocutors to argue with. On the one hand, I owe arguments against internalists and externalists who think no considerations are reasons in virtue of our wills. On the other hand, I owe arguments against purebred voluntarists who think that all reasons are. This chapter takes up the latter challenge. Since it's the most influential, I'll focus on Korsgaard's version of purebred voluntarism.

### 1. Korsgaard's Motivations

It'll help to start by putting the issues raised in Chapter 1 in Korsgaard's terms. On her view, we make considerations reasons for us by adopting "practical identities," which are descriptions under which we value our lives and find them worth living. My having a particular practical identity then requires that I treat certain relevant considerations as reasons.[59] For example, my willing that I have the practical identity of "cooking enthusiast" makes considerations about what would improve my cooking reasons for me. If I have this practical identity, I will treat the fact that some recipe employs a technique that's new to me as a reason to try it, when, absent this practical identity, this consideration would not have this normative significance (if I'm not an adventurous type, it might even count against trying the recipe).[60] A

59 For a full statement of this view, see Korsgaard (2009).

60 Note that I need not be conscious of all of these facts at the time of adopting the practical identity. It may turn out that being a cooking enthusiast requires honing my knife skills, even if this isn't obvious to me when I first take up cooking. Still, having the practical identity makes the relevant facts reason-providing when the circumstances arise.

large part of Korsgaard's project is to reconcile this picture with the claim that moral standards provide reasons for all agents. She does so by trying to show that a commitment to morality falls out of an analysis of agency. If so, anybody who can adopt practical identities at all turns out to adopt a practical identity that makes moral standards normative for her.

Korsgaard's primary motivation for accepting this picture is that she thinks it alone is in a position to explain the authority of reasons, or how they "get a grip" on us.[61] To unpack this a little, imagine that you are choosing between two pretty appealing options, and correspondingly you feel two pretty strong motivational tugs. Perhaps you feel both excitement at the possibility of introducing yourself to someone you admire, and nervous about coming off awkwardly, and so you are torn between saying hello and keeping to yourself. One familiar description of the phenomenology of these kinds of situations is that there are two parts of you, each bidding for control. One might then think that deciding what to do is a matter of adjudicating this dispute, either by deciding that one impulse isn't really part of you or by finding some way of reconciling them, so that you can put your full self behind one option or the other.[62]

Cases like these suggest that the question of what we have reason to do comes up for us because we are capable of getting some reflective distance from our motivations. We can experience them and then ask: is that kind of impulse one I can get behind?[63] Viewed from this

61 See Korsgaard (2009), p2.

62 This way of thinking about deliberation is pretty idealized. One worry is that it underestimates the reasons-responsiveness of our subpersonal processes. Railton (2014) provides some cases where our intuitions might track important facts even when reflection misses them, and makes some suggestions about how to understand the normative epistemology here. Another worry is that it might be a mistake to think of your whole self with what you can reflectively endorse; something can be a real part of me without my liking or even knowing about it. Connelly (ms) develops this point. I'm sympathetic to both of these worries, but I think we can safely set them aside here and interpret Korsgaard as abstracting away from them to focus on cases where reflection is going as well as it can.

63 See for instance Korsgaard (2009) p125-6.

angle, the question about whether some consideration is a reason turns out to be about whether that consideration is the sort for which we can close this reflective gap. As Korsgaard sees it, the chief advantage of explaining the normative force of an agent's reasons in terms of that agent's practical identities is that if an agent asks why some impulse has authority for her, voluntarists can answer in a way that appeals to deep facts about her, facts that she herself embraces. "Acting on that sort of consideration is partly constitutive of who you are," they can say. Korsgaard imagines source internalists having to say something like "Because of the desires you happen to have" and source externalists having to say something like "Because that's just what your reasons require."[64] *Prima facie*, the voluntarist answer seems more helpful in closing the reflective gap for an agent who entertains this question.

This point about voluntarists' answer to questions about the authority of reasons parallels one about the phenomenology of acting for reasons. Failing to introduce yourself – if that's what you end up thinking you have most reason to do – might feel like letting yourself down, or not being the person you think you are. So, whatever we end up saying about the authority of reasons, voluntarism has at least this much going for it: it seems to explain why some reasons feel weighty for the agents who have them, or why deliberating about what to do often feels like deliberating about what kind of person you want to be.

Korsgaard seems to think of this as a general point about what all deliberation is like. But, a similar argument might be developed in favor of views according to which only a subset of our reasons are self-created. Such an argument would say that voluntarism explains why *those* reasons feel weighty, or why failing to act on *those* reasons feels like letting oneself down. Just

64 Korsgaard (2009) p6

to report my own phenomenology, some kinds of practical failures feel more personal than others. For instance, failing to respond to ordinary, agent-neutral prudential reasons (say, to get enough sleep) doesn't bug me the way that failing to respond to reasons related to personal projects (say, to read some non-philosophy before bed) does, even when my reasons to get some sleep are about as strong as my reasons to read. One might think that voluntarism can help explain why these kinds of reasons feel different, even if one isn't a voluntarist about all reasons.

As an exegetical matter, it's a little hard to tell how literally to take Korsgaard's idea of practical identities making an agent "who she is." On the one hand, it seems clear that I could come to value my life under a quite different description, but still be the same person in the numerical sense that interests philosophers working on the metaphysics of identity. On the other hand – as we'll see when we explore her arguments for moral rationalism – Korsgaard puts a great deal of weight on the idea that if you lack a coherent practical identity, then there is really no "you" there at all, just a heap of subpersonal parts.[65] My best interpretative guess is that she'd want to say that having *a* practical identity is what's required for being a person, and that what depends on the continuity of your particular practical identity is something more like self-respect or integrity, rather than literal survival.

In any case, the first important question we need to ask in assessing Korsgaard's voluntarism is: does voluntarism really give us a uniquely good account of the authority of reasons?

65 See Korsgaard (2009) p165-170

## 2. Enoch's Shmagency Objection

Korsgaard wants to answer questions about the authority of reasons by saying something like: "acting on that sort of consideration is partly constitutive of who you are." We can make some progress in evaluating this response if we note that less abstract versions of this kind of explanation are familiar from talk about games and social conventions. Why not castle out of check? Because not doing so is part of what it is to play chess. Why shake hands instead of kiss cheeks? Because that's how you greet someone around here.

Korsgaard thinks that these kinds of explanations – explanations in terms of what's constitutive of some game, practice, activity, institution, and so on – are the key to responding to skeptical challenges.[66] They work by identifying an inconsistency in the skeptic: if something is required in order to count as participating in a given practice, then no one can consistently claim both to be involved in the activity and to be indifferent to the requirement. They are either not really involved or involved only defectively.

In "Agency, Shmagency: Why Normativity Won't Come from What's Constitutive of Agency" David Enoch argues that appeals to constitutive features of practices never really make any anti-skeptical progress.[67] Consider how the argument might go in a case like chess. Suppose that I am moving my rook around the board diagonally. "If you do that," you caution me, "you're not really playing chess." In response, it seems open to me to say "Fine, I don't care about whether I'm playing chess. I'm playing a different game where rooks move diagonally."

66 Korsgaard (2009), p28-29

67 Enoch (2006). Note that this is just one way to object. One might grant Korsgaard that if there is some norm constitutive of agency, it would explain the normative authority of reasons, but go on to wonder whether there really are any constitutive norms of agency, or whether they're contentful enough to give us a plausible account of the reasons everyone has (e.g. reasons to be prudent or reasons to be moral). I'll come to this other kind of objection below.

This example seems to show that constitutive claims don't necessarily give us reasons to engage in particular practices; they just tell us how to be involved in some practice if we antecedently have reason to be. Your response would have more teeth if I had some reason to play chess (instead of my weird diagonal-rook game), such that telling me I'm failing to play chess would mean I'm failing to respond to that reason. But if the explanation goes that way, it won't be just the constitutive rules of chess that explain why I have reason not to move my rook diagonally; it'll be them plus whatever reasons I have to play chess in the first place.

Enoch applies the same kind of reasoning to Korsgaard's reply to a skeptic about whether some purported-reason has authority for her. Just as I might not care whether the way I'm moving pieces counts as chess, I might not care whether the way I'm moving my body counts as agency (as Enoch puts it, I might be content with being a shmagent – someone like an agent but for whatever constitutive feature of agency I'm running afoul of). If the analogy holds, finding out that regarding something as a reason is required by my practical identity (as an agent, or some more specific identity I've taken on) shouldn't move me unless I already have a reason to be an agent or to have that more specific practical identity. But to give a reason to do those things would be to give up on the project of explaining normative authority in terms of those identities, because we'd have to appeal to a normative notion in order to get the explanation up and running.[68]

68 This analogy might seem a little fishy. Is agency really a practice in the same way that chess is? In a way, this is what the main response to Enoch in the literature turns on, and I'll come to it shortly. Nonetheless, for the sake of putting Enoch's objection most charitably, it's worth noting that nothing really hangs on the notion of a "practice" here. We might instead put the worry in terms of the activity of agency, or the game of giving and asking for reasons, or the shared project of acting for reasons, or whatever. Enoch, I imagine, would be happy to ask why we should participate in any of these things, so we don't need to be particularly precise about what a "practice" amounts to in order to formulate his worry.

This objection doesn't depend on anything in particular Korsgaard says about exactly which requirements are constitutive of agency, or what's constitutive of any particular practical identity. What generates the problem is that, seemingly for any practice, it's possible to ask for a reason to be engaged in that practice. As a result, Enoch's objection generalizes to other versions of voluntarism. Anyone who appeals to something like an agent's practical identity to explain the reasons she has will encounter the same kind of objection. So, even though I think Korsgaard's view has some serious problems, I need an answer to this objection just as much as she does.

In exploring how one might reply to Enoch's objection, I'm going to focus specifically on whether Korsgaard's strategy can explain the authority of reasons by appeal to claims about what's constitutive of agency, rather than any more specific practical identity. That's because I think that shmagency-like worries about whether it matters that something is constitutive of a particular practical identity (why commit to being a chess player and not a schmess player?) quickly reduce to questions about agency anyway. When pressed about why one has reason to do what one's practical identity requires one to do, Korsgaard can say that adopting and acting on these more specific practical identities is just how you go about being an agent, so long as you do so in a way that's consistent with any other requirements there are. If so, the reason we have to act on our (particular) practical identities is that that's how agency (in general) works. So, if there's an answer to the shmagency objection in the case of claims about what's constitutive of agency, it can be extended to answer shmagency-like objections in the case of more specific practical identities.

Since the shmagency objection arises because of an analogy between Korsgaard's

constitutive claims about agency and constitutive claims about other practices, the trick to replying to it is to find a relevant difference between agency and other practices that promises to explain why constitutive claims about agency give us reasons while constitutive claims about other practices don't. One candidate difference is that agency is not optional for us in the way that other practices are. According to Korsgaard, for example, agency is forced on us by the reflective distance we can have from our desires. Once we distinguish between the intensity and the authority of our desires, we can't help but use some principle to decide which ones to act on (even if that principle is a wishy-washy one, like to always treat our strongest current desire as authoritative).[69] The same isn't true for chess.

Unfortunately, Korsgaard leaves it open what exactly she means when she says that agency is non-optional. And we need to be careful here, because while there's definitely something plausible about this idea, on some readings of this claim it's going to come out false. We escape agency whenever we go to sleep, and even if sleep is the kind of thing we fall into rather than start intentionally, we might escape agency on purpose by plugging into the right kind of experience machine (one where we're not given subsequent choices about what to experience). So, more needs to be said to make Korsgaard's strategy work.

Luca Ferrero offers an interpretation of Korsgaard's claim that's intended to avoid these kinds of counterexamples. On his reading, the relevant difference between agency and other practices with constitutive requirements is that agency is the practice with the widest possible scope or jurisdiction. In other words, all other practices are entered or exited from within

69 Korsgaard (2009), p72-76

agency.[70] So, while I can choose not to play chess while remaining an agent, I can't choose not to be an agent while playing chess (since chess involves practical deliberation about which moves to make). Ferrero argues that the upshot of this observation is that no one can seriously entertain the question Enoch poses when raising the shmagency objection: should I be an agent or a shmagent? One would have to decide that question as an agent – as the kind of creature who decides things based on reasons – so to consider the question is already to be committed to the practice of agency. And if one is already committed to agency, then one can't be indifferent to its constitutive requirements.[71] So, being a shmagent just isn't a coherent option for us. Velleman seems to be thinking along these same lines when he says we can't make something an option for us just by appending prefixes to words.[72]

I think that Ferrero and Velleman are right that the inescapability of agency shows that someone who asks for a reason to be an agent rather than a shmagent is being inconsistent. Someone who doubts whether some consideration has authority for her can be answered by appeal to the commitments that make her an agent, and she can't be indifferent to those while still being in a position to ask for reasons. So, if this is what the shmagency objection was meant to deny, it fails. Nonetheless, one might think that even if we grant that Korsgaard can show that a skeptic about the authority of reasons is somehow inconsistent, this result still doesn't help us explain the authority of reasons. And doing that was supposed to be the main selling point of Korsgaard's voluntarism.

---

70 Ferrero (2009a)
71 Ferrero (2009a)
72 Velleman (2009), p143. I hedge here because Velleman's treatment of this question involves a few different lines of argument, and it's not obvious how much weight he wants to put on each. Enoch (2011b) helpfully disambiguates some lines of reply.

In a follow-up paper, "Shmagency Revisited," Enoch develops exactly this line of reply. He provisionally grants the idea that agency is inescapable in Ferrero's sense, but denies that this has any implications for whether or not it makes sense to ask whether to be a shmagent.[73] The key to his argument is the thought that a demand for justification can be forceful even if no one is in a position to coherently voice it. If so, Ferrero would be wrong to move from the premise that no one can consistently ask for reasons to be an agent rather than a shmagent to the conclusion that Korsgaard's explanation of the authority of reasons goes through.

Enoch offers two examples to get us on board with this thought. First, there's the Unwilling Patriot. This person sees no reason to engage in patriotic activities (singing the national anthem, rooting for her country in the Olympics, and so on), and sees plenty of reason not to engage in them (perhaps she's worried that they contribute to a dangerous kind of uncritical nationalism). Nonetheless, when opportunities to sing or root come her way, she finds herself singing and rooting, despite her judgment that she is being irrational. The Unwilling Patriot cannot coherently voice skepticism about whether the demands of patriotism have authority for her – after all, she remains committed to patriotism despite herself – but this doesn't seem to show anything about whether her skepticism is justified.[74] Second, there's the Paper Skeptic. This person thinks that philosophers shouldn't write journal-length papers. Perhaps she thinks that focusing on papers leads us to make narrow, logic-chopping moves in well-trod debates and that we could avoid this by working on book-length projects instead. It would be inconsistent for the Paper Skeptic to write up her reasons for paper-skepticism in a paper, but

73 Enoch (2011b)
74 Enoch (2011b) p15-16

whether we should take her arguments seriously doesn't seem to depend on whether she writes them up in the form of a paper or the form of a book.[75]

Enoch imagines someone wondering whether to be an agent or a shmagent as in much the same position as these characters. Entertaining the question requires some inconsistency, but what we are interested in is an answer to the question, not an evaluation of the person asking it. If so, something of the shmagency worry survives. We may necessarily be agents, but it needn't immediately follow that the constitutive requirements of agency are authoritative for us. Perhaps we should all be alienated participants in agency.

Enoch is probably right that these examples show something important: not just any inconsistency defeats a demand for justification. Nonetheless, I think there's room to wonder about how much more than that they show. Perhaps it matters that the Paper Skeptic *could have* expressed her arguments without inconsistency. Or perhaps it matters that the Unwilling Patriot's inconsistency depends on her particular psychology, while the inconsistency of asking for a reason to be an agent rather than a shmagent is more modally robust. It's a difficult question when identifying an inconsistency counts as anti-skeptical progress, and I don't have a general answer to offer here.

Nonetheless, even if we set the general epistemic question aside, I think it's plausible that, minimally, identifying inconsistency in a skeptic is progress when the inconsistency is a conceptual one. If, for example, I can't raise an objection to some philosophical view without relying on the premise that there are square circles, this really does seem like doom for my objection. Moreover, I suspect that this is the kind of inconsistency that Velleman thinks Enoch

75 Enoch (2011b) p21-22

faces. If you ask what reason there is to be an agent rather than a shmagent, Velleman says, you "aren't owed an answer because you haven't asked a question."[76] At any rate, this looks like the right kind of thing to avoid Enoch's worries about the Unwilling Patriot and the Paper Skeptic.

Enoch anticipates this line of reply. His response is just to deny that there is anything at all puzzling about asking for a reason to be an agent. If there isn't, replying to the shmagency objection this way looks like foot-stomping rather than finding a principled difference between asking for a reason to be an agent and asking for a reason to play chess.[77] So, as Enoch sees it, the debate about the shmagency objection comes down to whether or not one thinks that questions like "Why be an agent rather than a shmagent?" make sense.

This might look like a stalemate. Some philosophers (like Velleman) find questions like this fishy-sounding, while others (like Enoch) don't. If this comes down to a difference in intuitions, then we haven't made much progress. Enoch, however, isn't so concessive. As he sees it, this situation is only a stalemate if it's genuinely up for grabs whether questions like "Why be an agent rather than a shmagent?" make sense. Instead, he thinks, there is a strong presumptive case that they do. He reasons that if the Ferrero/Velleman line is correct, these questions are missing an argument-place. To make sense, they'd need to specify that one is asking from within agency. Read this way, they can be answered by appeal to the questioner's commitment to whatever's constitutive of agency. But, Enoch points out, these questions *don't* have the ring of questions that are missing an argument-place. They don't, for example, sound like "Is the Empire State Building taller?" If so, absent an argument for thinking that these appearances are

76 Velleman (2009) p144
77 Enoch (2011b) p31-32

misleading, we should reject the Ferrero/Velleman line and there's no stalemate after all.[78]

It seems to me that Enoch is being too hasty here. Even if questions like these don't cry out for a missing argument-place, it's not clear that missing argument places must be *obviously* missing. The answers to questions about weight, for example, depend on which planet they're asked on, but that's not obvious from the grammar of the questions.[79] And even if these sentences aren't fishy in this way, I think they do stand out as odd in other ways that suggest something may be amiss. Consider a question like "Why reason by *modus ponens*?" Like "Why be an agent," this question doesn't seem ungrammatical. But nonetheless, there is something distinctively fishy about it. If we don't reason by *modus ponens*, we end up like C.S. Lewis' tortoise: not reasoning at all.[80] If so, Enoch's focus on *grammatical* fishiness is distracting. Once our focus is shifted to fishiness *in general,* it's hard (at least for me) to share Enoch's feeling that "Why be an agent?" obviously makes sense.

Let's take stock of the dialectical situation. If I'm right that there really is a stalemate here, the shmagency objection doesn't raise any special difficulty for voluntarism. Though it looks devastating at first, it's hard to mount the objection without appealing to the falsity of the view it's meant to cause trouble for, and it's similarly hard to defend against it without appealing to the truth of that view. As a result, there's bad news for both Enoch and Korsgaard. Enoch hasn't non-question-beggingly shown that voluntarism is a non-starter, but Korsgaard hasn't non-question-

78 Enoch (2011b), p31-32

79 Moreover, I think Gricean norms provide a relatively straightforward explanation of why we don't expect a linguistic flag for the fact that we ask questions about reasons from within agency. If Velleman is right that there's nowhere else to ask from, then adding that we're asking from within agency would not be informative. So, it would be no surprise that speakers who hope to communicate information don't mention it. Thanks to Kathleen Connelly for pointing this out to me. See Grice (1975) for the the idea that contributions to a conversation aim at (among other things) being informative.

80 See Railton (1997) and Dreier (1997) for an extended discussion of this analogy.

beggingly shown that voluntarism has a uniquely good explanation of the authority of reasons.

One might think that the bad news for Korsgaard is worse. After all, explaining what makes reasons authoritative for us was supposed to be the main selling point of her view. Nonetheless, what's bad news for Korsgaard needn't be bad news for voluntarism. As we'll see in Chapters 5 and 6, voluntarism (at least as part of a hybrid view) has plenty more to recommend it, even if it doesn't have anything uniquely compelling to say about authority. Moreover, the bad news for Enoch is good news for hybrid voluntarists: we can appeal to agents' practical identities in order to explain the force of some reasons without fear of challenges from shmagents.

## 3. Doubts about Korsgaard's Rationalism

Leaving our discussion of shmagency in a stalemate may feel unsatisfying; we may seem not to have gotten to the heart of what's really challenging about the shmagency objection. In a way, I think that's right. I suspect that what makes the shmagency objection so intuitively appealing doesn't really have anything to do with abstract concerns about what kinds of arguments are permissible replies to skeptical challenges. The idea that agency is inescapable in Ferrero's sense sounds pretty plausible when we're just thinking about the bare concept of agency, something like “being a creature that does things on the basis of reasons.” Sure, *that's* inescapable. But if we substitute in a particular conception of agency (like Korsgaard's), agency suddenly looks much more optional. It'd be a big surprise if accepting the authority of the categorical imperative were inescapable. This suggests that what really motivates the shmagency objection is a different kind of concern, one about whether Korsgaard has a plausible conception of agency, instead of whether we can get normativity out of *any* conception of agency. So, we

need to turn our attention to that.

According to Korsgaard, being an agent is a matter of making considerations into reasons by having a practical identity, and having any particular practical identity implicitly commits you to having a moral one. More specifically, she thinks that all agents are committed to governing their behavior in a distinctively Kantian way: by only treating considerations as reasons when any rational agent could do so as well in the same circumstances. As a result, behavior that violates this constraint is either not action, or defective action.

One common reaction to Korsgaard is to say that it's just intuitively obvious that action doesn't constitutively require living up to Kantian ideals. The jumping off point for many of these kinds of objections is G.A. Cohen's case of the committed Mafioso.[81] The Mafioso is indifferent to morality (she won't hesitate to make her rivals sleep with the fishes), but, we are to imagine, this is compatible with her fully exercising her agency. She's making a moral mistake, but not an agential one.

There are a number of variations on this theme. Allan Gibbard finds it obvious that an Achaean warrior, who values himself as unflinching in battle, is both fully exercising his rational agency and is not thereby committed to morality. Sharon Street makes analogous claims about Perfectly Consistent Caligula, who aims to harm others and has no other ends that conflict with this (in theory or in practice).[82] For similar reasons, Chang is skeptical that Korsgaard could move from purely structural constraints on what we can will to substantive moral prohibitions, since characters like these are – by hypothesis – structurally rational.[83]

81 Cohen (2014)
82 See Gibbard (1999) and Street (2009). Similar arguments appear in Scanlon (2014)
83 Chang (2013b)

These examples focus on the relationship between agency and morality, but I don't think morality is really central to the force of the objection. What matters is that there is some set of standards we think all agents ought to conform to, and yet it seems possible to violate those standards without one's agency being defective. The standards of morality are an intuitively good candidate for this, but they might not be the only one. We might consider a committed daredevil who is indifferent to the norms of prudence, or a committed ascetic who is indifferent to aesthetic value. The point is: if you think there are any norms to which all agents have reasons to conform, there is room to wonder whether there are practical identities agents could adopt that do not commit them to following those norms. If so, it will seem to you that the ground of those reasons could not be the adoption of practical identities.

In any case, many philosophers take cases like these to be decisive. Ultimately, I agree that cases like these do give rise to extensional problems for Korsgaard's view. Nonetheless, I think it's too hasty to conclude that they do just based on confidence that the characters in these examples are fully exercising their rational agency. The problem is that “rational agency” is at least partly a term of art. Our intuitions about what rationality requires are likely to be influenced in part by the theoretical purposes we have in mind when we use that term. For instance, in economics it's common to assume a self-interest-maximization conception of rationality. Doing so makes it possible to develop elegant and explanatorily powerful models. And if one spends enough time working on and with those models, it will come to seem intuitive that rationality just is self-interest-maximization, but that doesn't follow from the usefulness of conceiving of it that way. Similarly, philosophers often use the idea of “structural rationality” to pick out agents

who are such that if they're making a mistake, they are not making a mistake by their own lights. This is an interesting category of agents, and it's worth having a name for them. But, that we chose a variant on "rational" as the name shouldn't lead us to conclude that structural rationality is all there is to rationality. If we find it intuitive that it is, we should at least entertain the possibility that we have this intuition because we metaethicists have thought about structural rationality a lot, and that a competent user of the concept "rationality" who did not travel in the same philosophical circles might not. So, I propose, we should be humble about how much our intuitions have to say here.[84]

Things are especially unclear when we note that Korsgaard can allow that characters like these are exercising their agency, just defectively. Sure, Korsgaard can say, we can be confident that Perfectly Consistent Caligula is an agent, but a *full-blooded* one? While it may be obvious that the characters Gibbard and Street imagine are agents, it's not obvious that they are non-defective agents. If they are, we need some argument for that.[85]

Moreover, to the extent that it is intuitively clear what to say about cases like these, it's not clear that we should treat those intuitions as fixed points. Many philosophers find it hard to believe that agency has a constitutive aim, but many philosophers also find it hard to believe that there are mind-independent normative facts. Nonetheless, source externalism is a respectable

84 This point is easier to appreciate, I think, if we briefly ignore questions about the relationship between morality and rationality. It's much easier to entertain the idea that rationality requires prudence, for example. But once we entertain that idea, we've agreed that rationality full stop might require more than structural rationality.

85 The intuition that someone like Perfectly Consistent Caligula is a full-blooded agent seems to depend on the idea that being an agent needn't involve recognizing morality. This might be right, but I don't think it's pre-theoretically obvious. After all, it's plausible that being an agent involves responsiveness to reasons. One might think, then, that being a full-blooded agent requires being responsive to the full range of reasons there are. And it's plausible that moral reasons are some of those. I don't mean this to be an objection to moral anti-rationalists (it's obviously question-begging against them), my point is just that it's just as much question-begging against Korsgaard to start with an anti-rationalist conception of agency.

view; we should believe in such facts if they turn out to be the best way to explain things we want to explain. I don't see why these surprising claims about ontology should be any different from Korsgaard's surprising claims about agency. That something is initially hard to believe isn't a decisive objection against a philosophical view.

The upshot is that apparent counterexamples to Korsgaard can't do the argumentative work all on their own. If we're going to make good on the idea that Korsgaard's voluntarism has problems with extensional adequacy, we need to engage more directly with her arguments that there are moral commitments that are constitutive of agency. Coupled with a reason not to buy those arguments, cases like the Achaean warrior can spell extensional trouble for Korsgaard. Without such an accompanying reason, at best such cases show that the conclusions of Korsgaard's arguments are surprising.

So, we need to turn our attention to those arguments. Here's my best attempt to reconstruct them. Korsgaard's argument that we're each committed to roughly Kantian ideals starts with a common idea from action theory. Part of what distinguishes things we do from things that merely happen to us is that when we do something, we can see ourselves as the cause of what happens. This is why, Korsgaard thinks, we are committed to the instrumental principle.[86] Doing something for the sake of some end (rather than merely being around when something brings about that end) is a matter of being the cause of that end's coming about. And being the cause of that end's coming about is a matter of taking the means to it. So, anyone who acts to bring about ends is committed to taking the means to them.

Korsgaard thinks that this same observation about distinguishing what we do from what

86 See Korsgaard (1997)

happens has another consequence. It requires us to identify ourselves not with any of our particular motivational impulses, but instead with something over-and-above those impulses. Otherwise, it wouldn't be *me* deciding which impulse to treat as authoritative.[87] Once we acknowledge this, she thinks, we're in a position to ask how we're supposed to decide which motivational impulses to grant authority to.[88] Her argument for moral rationalism proceeds by introducing two constraints on this decision.

The first constraint is intertemporal. It says that when we make a decision about what will count as a reason for us now, we need also to provisionally count it as a reason when similar circumstances arise later, unless there is some good reason why not.[89] So, for example, this constraint requires that if we decide that the prevention of future toothaches is a reason to go to the dentist now, it will also be one later, unless relevant circumstances change (e.g. the dentist is discovered to be incompetent). This clause about changing circumstances is important; new circumstances are often going to demand new decisions about what's authoritative for us, and this constraint doesn't say anything about how to do that. Nonetheless, this constraint does impose some stability on our practical identities. It prevents us from arbitrarily making and taking back commitments to treat classes of things as reasons.

Korsgaard's argument for imposing this constraint is that without it, we lose a grip on what merely happens in us and what we choose to do. Suppose that I'm choosing between two options, each appealing in its own right. If I'm to see myself as choosing between them, rather than passively letting my desires battle it out, I need to choose on the basis of some principle.

87 Korsgaard (2009), p1
88 Korsgaard (2009), p1
89 Korsgaard (2009), p72-76

The principle might be a substantive one (maybe there's some part of my practical identity that tells me what to do) or a wishy-washy one (maybe I decide to go with whatever desire I feel more intensely), but unless there's *some* principle or other it's hard to see where I come into the picture, rather than just my desires for the two options.[90] And for this choice to be based on a principle, rather than merely being arbitrary, the principle will have to be presumptively applicable to other, similar cases. That's what makes it a principle rather than a judgment about a particular case.[91] This needn't mean that the next time I'm choosing between these two options in these circumstances I have to choose the same way, just that if I choose differently, I'll need a reason for the change.

An agent who went around following this requirement would end up with a number of principles, each governing different kinds of choices with various degrees of generality. These principles coalesce to form the broader descriptions under which we value our lives (in Korsgaard's terms, our practical identities). So, for example, consistently deciding to treat certain kinds of impulses as authoritative and others as mere distractions or temptations can make me a baseball fan, an amateur cook, a beer nerd, or someone's friend.

This raises the possibility that these high-level commitments can come into conflict. Sometimes being a friend requires helping to load the moving van, even when you could have gone to the ballgame. So, agents need a way of ordering and unifying their principles. This gives rise to Korsgaard's second requirement on decisions about which motivational impulses to treat as reasons: agents must do so in a way that preserves the unity of their practical identity.[92]

90 Korsgaard (2009), p72-76
91 Korsgaard (2009), p72-76
92 Korsgaard (2009). See especially Chapter 8.

To see what this principle amounts to, consider some ways of organizing your principles that are prone to breaking down. Suppose that you care both about doing what's morally right and about being appreciated for doing so, and you treat these as equally central parts of your practical identity. You might, unluckily, find yourself in circumstances where you cannot do both. In such a situation, either way you choose you'll have betrayed your deepest commitments, because your principles tell you both to embrace and reject the same things. Such an agent can act on her most basic commitments, but only contingently – only so long as she isn't unlucky enough to find herself in situations where they diverge.[93] Or suppose that you accept the wishy-washy principle mentioned earlier, and govern yourself by always treating your most salient current impulse as authoritative. As this kind of agent, you can distinguish what you really care about from what's merely a temptation, but only when the temptations are weak. When they get strong enough, they just become the thing you really care about.[94] In such a case, you're again in the position of betraying your fundamental commitments no matter what you do. Korsgaard's requirement that we preserve the unity of our practical identities essentially says that you need to avoid being the sort of person who is susceptible to this kind of bad luck. In other words, your ability to function practically cannot be this contingent; you need a set of commitments that can guide you come what may.

This kind of contingency might not seem like a big deal. After all, in many cases our deepest commitments don't tear us apart like this, so it seems odd to think that the possibility that they might shows that we are doing something wrong in the cases where they don't. Korsgaard

93 Korsgaard (2009), p165-170
94 Korsgaard (2009), p165-170

isn't perfectly clear about this, but I suspect that her worry traces back to the goal of identifying something over-and-above an agent's motivational impulses and giving this thing a role in the story of action. If our practical identities are such that they can be split against themselves in the way these cases illustrate, they can't be used to identify one agent behind an action. Instead, it'll just be different sub-agential parts battling it out, and that's no more unified than different desires battling it out. So, even when we're lucky and circumstances don't test our integrity like this, it's nonetheless true that there's not a unified agent behind each choice. That's why our practical identities need to be counterfactually stable, rather than just stable as things are.

Putting these two requirements together, we need to decide which of our motivational impulses to act on in a way that has presumptive authority going forward, and that is sufficiently counterfactually robust so as not to put us at risk of self-betrayal. If we accept both of these requirements, it's a short step to Kantian morality. Because we don't know in advance what situations we'll encounter and how our principles might come into conflict, the way to satisfy the second requirement is to choose principles *any rational agent* could accept. This is because, as Korsgaard puts it, circumstances could conspire to place us in anybody's shoes.[95] And, by the first requirement, once we have structured our practical identities this way, we can't shed them arbitrarily. So, the particular commitments that make us into unique agents have to be structured and constrained by the moral ones that make unified agency possible.

While Korsgaard's goal is to find actions among things that happen and agents among sub-agential impulses, I think there are other philosophical resources that can be marshaled in support of her two constraints. Consider the intertemporal constraint and the literature on

95 Korsgaard (2009), p214

diachronic agency.

As Ferrero points out, there are more and less sophisticated ways an agent might go about working with her future selves.[96] On the simplest model, there needn't be any real coordination between different time-slices. Ferrero's example is a bacterium moving toward a glucose source. Each time-slice of the bacterium just needs to move toward the highest adjacent density of glucose. The cumulative result of this is movement along a path toward the food source, but no stage of the bacterium need have any conception of what it's doing or what's required of the other stages.[97] The cumulative movement just emerges from the independent moves of each time-slice. Somewhat more sophisticated agents can conceive of future goals, but go about pursuing them such that each stage of the agent behaves independently of the other stages.[98] For example, suppose that I want to drive to Colorado, and that there are two roads that will take me there. Once I choose a road, I needn't ever revisit this choice or let it inform my future choices; I just need to pay attention to local driving conditions and I'll end up in Colorado. In a case like this, each stage of me needs to be in the loop about what the end goal is (I can't just stop driving and hang out in Las Vegas), but no further coordination is required. I can go about each stage of the drive pretty much independently of the others. Finally, there are cases where success requires genuine coordination between temporal parts. Ferrero's primary example of this is rational discourse.[99] If I am trying to convince you of something, each temporal stage of me can't proceed mechanically from the situation left to it by the previous stage. Instead, each stage needs to know what past stages were up to (e.g. that I introduced an assumption for reductio rather than for

96 Ferrero (2009b) p408
97 Ferrero (2009b) p408
98 Ferrero (2009b) p408
99 Ferrero (2009b) p415-419

conditional proof) and needs to be able to react to new information in a way that's structured by the moves made by earlier stages (e.g. revising premises in light of objections) and by my ultimate goal (e.g. trying to revise those premises in a way that preserves their usefulness for my argument).

Imagine trying to do something that required this third kind of diachronic agency if you didn't respect Korsgaard's intertemporal constraint. If decisions about what to pursue didn't even have *pro tanto* authority going forward, no time-slice of you could rely on any future one to take the next steps in the plan, or to be concerned for how this plan interacts with your others. Instead, you'd have to re-adjudicate the question each time, and there'd be no genuine collaboration between time-slices; when one does what the previous one planned or the next one needs, it'd just be good luck.

These distinctions within diachronic agency lend support to Korsgaard's idea that when we fail to act in accordance with the intertemporal constraint, we're falling short of the constitutive standards of agency. If we don't respect the constraint, there are kinds of agency that just aren't available to us. We might still be agents, but less than fully so. And just as there are different kinds of collaboration between time-slices of an agent, there are analogously different kinds of collaboration between agents.[100] So, similar things can be said in favor of respecting Korsgaard's second constraint; it makes possible more sophisticated kinds of collaboration between agents.

In short, I think that arguments like Korsgaard's actually do go some way toward

100For some further discussion of this idea, and how it might be leveraged in a non-Korsgaardian argument for moral rationalism, see Brink (1997b).

establishing moral constraints on action. But, they don't go all the way. The problem is that one can respect both constraints (intertemporal and interpersonal) without thereby committing to acting in a way that any rational agent can share. Integrity requires some counterfactual stability, but not this much.

David Sussman develops this objection in the following way. Consider the sorts of commitments that most thoroughly structure our sense of what's possible for us and the ways we can relate to others, like being a parent or a devoted follower of a particular religion. Sincere commitment to these kinds of things often forecloses other possibilities; they put limits on what we can imagine ourselves changing and still being ourselves afterward. But if this is so, maintaining a sense of oneself as unified agent can't require that sense to be stable in anyone's shoes. Otherwise, these kinds of deep commitments would be impossible; part of going in for them is not being able to see (all of) yourself in the shoes of people who don't.[101]

Sussman takes the upshot of these considerations that integrity doesn't require having a sense of yourself that will be stable come what may; it only requires stability in cases where one can imagine being the person in the circumstances.[102] So, for example, integrity requires that I have principles that won't disunify me when living up to them is inconvenient, but it doesn't require that I have principles that would help me face Sophie's Choice. The difference, on Sussman's view, is that I can imagine what it would be to live up to my principles, but not what it would be like for a parent to be in that situation.

I think Sussman is right that there's a problem for Korsgaard here, but I don't think it

101 See Sussman (2015)
102 Sussman (2015)

really has to do with what situations agents can imagine their way into. Often imagination is a flimsy guide to psychological possibility; we often don't know how we'll make sense of ourselves in some situation until we try it out.[103] So, Sussman's gloss on integrity makes integrity too easy: it fails to distinguish cases where something is hard to imagine because it's scary or new from cases where something is unthinkable because it's genuinely not an option for us given who we are. Nonetheless, so long as there are some cases where given who we are some other way of being is unthinkable, we have counterexamples to Korsgaard. Being a unified agent doesn't require that your most fundamental commitments be ones that any rational agent could share.

Korsgaard might want to reply that, as demanding as her constraint sounds, relaxing it in a Sussman-like way just amounts to admitting that we are not fully unified, and as a result there's no single entity behind our decisions, just some disorganized subpersonal parts. But this reply, it seems to me, misidentifies the kind of unity required for agency. It would be weird, I think, for a parent *not* to feel conflicted in a Sophie's Choice situation; even if one finds some reason for choosing one child over the other, properly appreciating the reasons one has in this situation (in other words, functioning well as an agent) seems to *require* feeling torn. Not feeling that way would amount to not really getting what makes the situation bad.[104] Being able to have a reflective take on your reasons needn't require having a comfortable or unambiguous take on them.

The upshot for Korsgaard is that her view faces serious extensional problems of just the

103 For some cases like this, and for an interesting attempt to draw some more general conclusions about practical reason from them, see Paul (2015).

104 See Coates (2017) and Gunnarsson (2014) for arguments along these lines.

kind that Cohen, Gibbard, and Street suggested. What if the thing I'm committed to, such that I wouldn't be me were I not to pursue it, is immoral? In cases like this, Korsgaard seems forced to say that we have reasons that intuitively we don't. Crucially, though, we now have an explanation for why we should believe such cases don't just beg the question against Korsgaard. They're compatible with her two constraints on full-blooded agency if we imagine these agents to have commitments that run deep enough to rule out being themselves without them.

This objection opens the door to another one. Even if Korsgaard succeeds in showing that everyone is committed to some moral requirements, it remains an open question what happens when these requirements come into conflict with the requirements brought in by that agent's other commitments. When these other commitments are ones the agent couldn't recognize herself without – i.e. when they're the kind Sussman has in mind – it looks like we have no particular reason for thinking that the moral requirements should have priority. In other words, we lack an argument for thinking that the constitutive requirements of agency always supersede or outweigh other requirements that an agent might take on. But if so, cases like the Mafioso, Perfectly Consistent Caligula, and the Achaean Warrior raise problems again. Even if we can show that Caligula has *some* reason not to torture, it still seems like a bad result if Caligula turns out to have *most* reason to torture.

## 4. Doubts about Bootstrapping

Sometimes, when philosophers are articulating the extensional worries we considered above, they advertise them as though they show that there is something implausible about the voluntarist idea itself, rather than just its implications about cases. For example, when Cohen

first introduces the Mafioso case, he suggests that it illustrates how if voluntarism were true, it would give us an implausible kind of control over our reasons.[105] Less specifically but more vividly, when considering the same type of case Hilary Kornblith says that the idea that we can make some consideration a reason by reflectively endorsing it as such "makes a mockery" of the idea of a reason.[106] This sounds like a deep worry – as we'll see in the next chapter, internalism also faces some extensional challenges, but it doesn't get philosophers' hackles up in the same way – so to finish our assessment of Korsgaard's voluntarism it's worth pausing over what might be driving anti-voluntarist intuitions like these.[107]

It might be useful to start our search by thinking about the religious analogue of the voluntarist views that concern me here. Historically, philosophers have objected to the proposal that we explain what's morally right in terms of what God would choose, endorse, or approve of on the grounds that such views make the content of morality *arbitrary*.[108] This arbitrariness might also give rise to a problem of authority. If God doesn't choose, endorse, or approve of some actions in virtue of their morally relevant features, it's not clear why we should care about his choices, endorsements, or approval. Similarly, if my choices determine what I have reason to do, this might seem objectionably arbitrary. I've adopted the practical identity of an amateur cook, but I could have chosen differently; I could have adopted the practical identity of a billiards shark or sketch comedy writer. Had I chosen those things, the secular voluntarist view says, I would have had different reasons.

105 Cohen (1996) p167
106 Kornblith (2012) p129-133
107 From presenting some of these ideas at conferences, my anecdotal sense is that Kornblith is not alone in having this reaction.
108 The dilemma Socrates presents Euthyphro with at the end of their discussion of piety is the jumping off point for this worry. See Plato, *Euthyphro.* Irwin (2008) discusses the historical trajectory of this objection.

This worry gets at something important, but I don't think arbitrariness can be the whole story. The problem is that there's no general connection between arbitrariness and a lack of normative authority. It's arbitrary that we drive on the right rather than the left in San Diego, but that's not a reason not to care about driving on the right. Similarly, even if it was arbitrary that I took up cooking rather than comedy, I did in fact take up cooking, and this is relevant to the reasons I now have. Voluntarists give a different explanation for this than do internalists and externalists, but all three explanations are compatible with some arbitrariness.

A better suggestion along these lines is that voluntarism makes *all* of our reasons arbitrary. We have genuine reasons to drive on the right hand side of the road, this objection might say, but that's only because there are non-arbitrary reasons not to put ourselves and others at risk. It's worth noting, however, that I don't think Korsgaard would accept this characterization of her view. It's true, she thinks, that all our of our reasons are a product of our wills. But it needn't follow from this that all of our reasons are arbitrary. Because of the constitutivist constraints she recognizes on agency, it's not true that we could have willed any which way. So, the success or failure of this worry about voluntarism really hinges on the success or failure of Korsgaard's attempt to address extensional objections. It may make extensional problems worse if Korsgaard has them (which, I argued above, I think she does), but it's not an independent reason to be skeptical about voluntarist views in general.[109]

Here's a related possibility, inspired by Cohen's description of the Mafioso case. Voluntarist views attempt to give a reductive explanation of normativity, i.e. they explain

109 I'll return to this issue in Chapter 7 to see if this worry gets traction against my preferred hybrid voluntarist view.

normativity in terms of something non-normative. But one might worry that something about the way this reduction goes leads us to lose a grip on the phenomenon we were trying to explain. For instance, it's central to the idea of a reason that it's something that counts in favor of an alternative, and so can guide us when we're thinking about what to do. Perhaps the voluntarist analysis of reasons undermines this idea. How, one might wonder, can a reason guide you or count in favor of something if it's always up to you to take back the commitment that makes it a reason?[110]

This kind of worry shares some features with "bootstrapping" objections in epistemology and the philosophy of action. In philosophy of action, some worry about whether forming an intention to do something can ever count in favor of doing that thing. If it did, this might give us an implausible kind of control over what we have most reason to do.[111] In epistemology, some worry about whether we can ever confirm a belief-forming mechanism's reliability by checking it against itself. For example, it might be inappropriate to find out whether the gas gauge in my car is accurately reading the tank by checking the gas gauge.[112] Voluntarist reasons might seem suspect for similar reasons.

This is an important question, but I'm not sure that it leaves Korsgaard with a distinctive problem. As with arbitrariness, there is no general connection between a reason's being contingent on a revisable decision and its normative authority. Any view about the source of normativity is going to end up saying that what we decide can have downstream effects on what

110 I'm not sure if this is the objection Cohen initially had in mind. Maybe it was an extensional worry, or maybe it was some other kind of suspicion about voluntarism. Nonetheless, this proposal seems worth exploring whether or not it's the one Cohen intended.

111 See Smith (2016) for discussion.

112 See Weisberg (2012) for discussion.

reasons we have. But if this is right, then all of the views are going to have to accept some contingency. Suppose I have a reason to go the gym today, and that I plan do it in the late morning. I could have decided this differently, and I could revise this plan; maybe I'll get a brunch invitation and decide to go earlier, or maybe I'll get in a writing groove and decide to postpone until the afternoon. Nonetheless, this openness to revision doesn't cast any doubt on whether, absent revision, I have reason to go to the gym in the late morning. This will be true whether you think that my reason to go is explained by my practical identity (say, that I've committed to some exercise regimen), my desires (say, for mental clarity), or the normative facts (say, that a good life requires cardiovascular health). Either the contingency voluntarism admits isn't distinctive – in which case there's no objection here – or it is – in which case whatever makes it distinctive is what's really driving this worry, rather than contingency or revisability.

If contingency itself isn't the problem, what is? I suspect that what's behind this line of thought is a worry about how *stable* our reasons are. If I were in a situation where it really were an open question whether it makes sense to go to the gym in the late morning, and if re-hashing this question kept getting in the way of actually getting out the door, then there really would be cause for concern about whether contingency undermines normativity. But notice that if we interpret the objection this way, the bootstrapping worry doesn't take us much further than the extensional objections. If Korsgaard were right that the nature of agency places some constraints on what we can rationally will that turn out to be quite substantive, there'd be no reason to worry about the stability of our wills. So, as we saw with the classic arbitrariness objection, this doesn't look like a problem with voluntarism per se, just a problem with Korsgaard's constitutivist

strategy for addressing objections about the extension of reasons.

Still, someone who shares Kornblith and Cohen's discomfort with voluntarism might think that these observations miss the point. Sure, voluntarist reasons can exhibit stability or provide practical guidance, they could grant, but not in the right way. What's missing, such an objector might think, is that reasons have *authority* for us. As Philippa Foot emphasized, something can be a stable practical guide without having authority; the rules of etiquette might consistently tell me the same thing about how to respond to dinner invitations, but this doesn't settle whether the rules of etiquette have authority over me.[113]

In evaluating this objection, we need to be careful about what we mean by "authority." For there to be a real problem for voluntarist views, there would need to be a sense of the term that's plausibly both (a) part of the concept of a reason and (b) not compatible with the voluntarist story about in virtue of what considerations count as reasons. So, we need to distinguish some different senses of authority and consider whether they satisfy both conjuncts.

One possibility is that when it comes to reasons, "authority" means something like "justification." Suppose that you tell me that etiquette requires that I not put my elbows on the dinner table. I might respond, "But this is comfortable. Why does etiquette have authority?" You might answer this question by appealing to reasons I have to conform to the demands of etiquette: maybe my elbows are dirty, or some of our guests care a lot about etiquette and I'll hurt their feelings by ignoring it. This sense of authority is clearly part of our concept of a reason; if reasons are anything, they're considerations that justify. But this way of understanding the objection clearly begs the question against voluntarism. The voluntarist thesis just is that some

113 Foot (1972)

considerations justify in virtue of our willing that they do. So, the objection can't just be a denial of that claim; to make it work, we need to identify something about authority or justification that's incompatible with voluntarism.[114]

Perhaps the kind of authority relevant here has to do with the legitimate changing of someone's normative status. Consider a ticket-taker at the movies. When you hand over your ticket, you are granted permission to enter the theater; if you don't have a ticket, you're not granted permission. This change in your normative status (permitted vs not-permitted) need not come along with any non-normative changes in the environment; it's possible for you walk in whether or not you're allowed. Nonetheless, the ticket-taker's authority determines whether you are entering with or without permission. This kind of authority clearly satisfies (a). At a minimum, our reasons determine whether our actions have the status of justified or unjustified, and perhaps other statuses that can be expressed in those terms (like, possibly, required or prohibited, prudent or imprudent, etc). Still, for there to be a problem for voluntarism here, this conception of authority would need to satisfy (b) as well. But if this kind of authority is incompatible with a voluntarist metaphysics, I don't see how. It seems perfectly coherent that a ticket-taker could give herself authority to see a particular movie, for example. So, there's no mismatch here between the voluntarist story and our idea of a reason.

A third way to understand authority is to think of it in terms of the legitimate exercise of coercive power. While anyone can tell you what to do, only someone with authority can make

114 Maybe some philosophers who express this kind of worry would be okay with the question-begging version. They might just think it's a conceptual truth that reasons are prior to the will rather than vice versa. Watson (2009) suggests something like this, drawing on Raz (1986). If so, I'm not sure what to say, other than that we shouldn't mistake confident assertion for conceptual truth. At any rate, philosophers who conceive of the objection this way shouldn't take it to be capable of convincing anyone who finds the voluntarist idea attractive. So, at best, this objection would get us to a stalemate of intuitions, not philosophical progress.

you do it without thereby wronging you. For example, any pedestrian can signal for you to stop your car, but only the police can arrest you if you don't. This way of thinking about authority certainly satisfies (b) above. If you can always take back the commitment that grounds some reason, then there's no one in a position to coerce you into complying with that reason. If someone were to try, you could just take back your commitment. But I'm very suspicious about whether this conception of authority satisfies (a). There are no Reasons Police that coerce us into doing what we have reason to do, so if this is what's required for authority then no view about the metaphysics of reasons is going to satisfy it. The authority reasons have over us must be different from the authority a queen has over her subjects.

Of course, this might not be an exhaustive list of conceptions of authority. So, there could be a kind of authority that, say, reasons with an externalist source have but reasons with a voluntarist source do not. Nonetheless, for there to be an objection here, one would have to specify what that sense is in a non-question-begging way (i.e. having authority can't just be a matter of having a non-voluntarist source) and make it clear that the to-be-specified sense of authority is required for being a reason. Absent such an argument, I don't think there's a serious worry about authority here, or at least not one that can be put in theory-neutral terms.

Joseph Raz offers another candidate interpretation of this objection. Raz worries that on a voluntarist view, we can't explain how anyone makes normative errors. If my reasons are a function of my will, it seems that any time I'm at risk of making a normative mistake – i.e. acting in a way that's not supported by my reasons – I can just change my will such that what I'm doing is supported by my reasons after all. If so, I seem to be immune to rational criticism.[115] If it were

115 Raz (1986)

true that voluntarism ruled out the possibility of normative errors, that really would be a troubling consequence of voluntarism, but I think Raz is making two mistakes here. First, voluntarists don't have to say that willing any old thing is rationally permissible; there might be constraints. For Korsgaard, these come in the form of constitutive claims about agency. So, here again, we seem to have a restatement of extensional worries about voluntarism rather than a further problem. Second, even if agents *could* rationally will anything at all, it wouldn't follow that they actually do so whenever they face potential criticism. I might make a commitment and then fail to live up to it without thereby abandoning the commitment. In fact, this seems to be the typical case; I don't give up on the goal of finishing my dissertation every time I procrastinate working on it.[116] The mere possibility that I could abandon my commitment doesn't show that I'm not criticizable when I in fact don't live up to commitments I still have.

Another way to develop the anti-voluntarist intuition might be to point to some facts about the phenomenology of deliberation. When we're thinking about what to do, we don't – or at least don't primarily – think about our practical identities. When I pour myself a cup of coffee, for instance, I think "a coffee would be nice" rather than "I'm the kind of person who endorses drinking coffee for its taste." So, perhaps what seems odd about voluntarism is that the explaining our reasons in terms of our practical identities Korsgaard is presenting an implausible account of what deliberation is like. Korsgaard certainly sometimes writes in a way the invites this interpretation, but I don't think voluntarists need to. Two points are worth making. First, as Schroeder points out,[117] what metaphysically explains why some consideration is a reason needn't

116 I'll have more to say about this kind of case in Chapter 6.
117 Schroeder (2007a) p24

be part of that reason. So, it could be that "coffee tastes good" is my reason for drinking it, even though the thing in virtue of which that's a reason is more self-directed. Second, deliberation is more of a mixed bag than this objection makes it out to be. In deliberation about deeply personal matters, like which careers to pursue, passions to follow, or religious activities to participate in, looking inward seems perfectly appropriate. So, I don't think the phenomenology of deliberation points univocally in a pro-voluntarist or anti-voluntarist direction. This provides some motivation for following Schroeder's lead and not assuming that our story about in virtue of what a consideration is a reason needs to be identical to our story about how we deliberate about reasons.

Here's one final suggestion: maybe the idea that we can create reasons at will looks suspicious because of the possibility that we might go about this willing badly. If voluntarism were true, this thought goes, we could create reasons even when we're wildly misinformed, or overtired, or blind drunk. And maybe that would be implausible, even setting aside extensional considerations (suppose, luckily enough, what I commit to while misinformed turns out to be what I would have committed to after doing my homework). But as with the previous proposals, this doesn't look like the kind of thing that could ground general skepticism about voluntarism. Korsgaard could just insist that only commitments undertaken under reasonably favorable conditions count.[118]

To sum up: I think that philosophers who have found voluntarism implausible really are on to something – purebred versions of voluntarism like Korsgaard's really do have trouble with cases like the Mafioso – but the problem here isn't a conceptual one about voluntarism. It's just

118 Smith (2016) makes a similar point about forming intentions.

that Korsgaard's rationalist arguments fall short, and so her view has a hard time getting the extension of reasons right. So, these cases don't show that voluntarism is a non-starter; they just show that a respectable version of it is going to have to have a way of finessing the relevant cases.

Of course, it might be that while Korsgaard's particular arguments fall short, there could be some other path from structural features of agency to moral requirements. If so, a purebred voluntarist could avail herself of that account. But, despite my best efforts, I can't think of one.[119] So, while I don't take the discussion here to have conclusively ruled out purebred voluntarism, I think it at least motivates looking for other ways to develop the voluntarist idea. That will be our task going forward.

119 I don't mean this to be glib. Korsgaard was my entry point into philosophy of action, and – at least as a matter of autobiography – talking myself into hybrid voluntarism was mostly a matter of talking myself out of a purebred view. I suspect that for most readers things will go the other way around.

# Chapter 3

## Problems for Purebred Internalism

In the last chapter, I argued that voluntarism is implausible as a purebred account of the metaphysics of reasons. From here, one might draw one of two conclusions. One possibility – the one I'll ultimately argue for – is that we should supplement a commitment to voluntarism with another view about the metaphysics of reasons that can help it address the extensional problems Korsgaard faces. Another possibility is that we should give up on voluntarism and go for a purebred version of externalism or internalism instead. My motivation for going the first way rather than second is partly that I think hybrid voluntarism offers some unique attractions, and I'll discuss those in Chapters 5 and 6. However, it's also partly that I think there are some important costs to the purebred versions of alternative theories. So, as scaffolding for the arguments to come, I'll now turn to explaining those costs. I'll start, in this chapter, with internalism.

### 1. Motivations for Internalism

When introducing possible views about the metaphysics of reasons in Chapter 1, I noted that internalism differs from externalism in that it offers a reductive analysis of reasons; it attempts to explain what makes some consideration a reason in non-normative terms. For some philosophers, this is a chief appeal of purebred internalism. If, for example, you find the idea of mind-independent normative facts metaphysically spooky or epistemologically suspect, internalism may sound appealing. Or, if you're suspicious about whether externalists can explain the connection between judging that you have reason to do something and being motivated to do

it, internalism's explanation of reasons in terms of desires might seem like a helpful resource.[120]

These motivations for internalism turn on objections to externalism, so I'll take them up when I offer a qualified defense of externalism in Chapter 4. Nonetheless, there are ways of motivating internalism on its own terms. First, one might draw on the observation we made in Chapter 1 about the constraints that different senses of "reason" put on each other. If there is a tight connection between justifying reasons and motivating reasons, and motivating reasons are plausibly explained in terms of the desires of the agents they motivate, then desires might look like a good place to start for explaining justifying reasons too. This thought might be developed in various ways. Perhaps, for instance, there is a "resonance constraint" on justifying reasons, that they be the sort of thing that could also be motivating reasons.[121] Alternatively, it might be that justifying reasons must be the sort of thing that it is appropriate to offer other people when trying to come to an agreement with an interlocutor, i.e. reasons that they can see the force of. This kind of constraint might also seem to rule out reasons that are in no way connected to an agent's pre-existing set of desires.[122]

Second, as Markovits points out, there is something appealingly *epistemically humble* about the internalist idea.[123] While a non-reductive view about the metaphysics of reasons must come paired with a view about what the normative facts say, internalists can be officially agnostic. They can say that what we have reason to do is what promotes the satisfaction of our

120 Williams (1979) is the standard jumping off point for arguments like these. Though see Shafer-Landau (2003) p181-2 for another interpretation of what Williams is up to.

121 See Brink (2008) for discussion of how we might interpret this constraint and whether those interpretations uniquely favor internalism.

122 See Manne (2014) for one way of developing this argument.

123 See Markovits (2014) p54-58. Though she doesn't express it this way, something like this idea also seems to motivate Street (2009).

desires, and let the particular content of our reasons fall out of that. Viewed from one angle, this looks like a manifestation of the idea of equal respect for every rational agent; each agent, by virtue of having desires, contributes to making the normative world what it is.[124]

Third, Schroeder points out that if we're interested in figuring out what makes a consideration a reason, it can be instructive to compare cases where that consideration is a reason for one agent and not for another.[125] This is an application of more general point about explanation. If we want to know in virtue of what some geometric figure counts as a square, we can't look only at squares. In order to learn, for example, that having sides of equal length is not sufficient, we'd also have to look at other shapes with this feature, like equilateral triangles. Once we make this kind of comparison, though, it looks like it's often a difference in desires that explains the difference in whether some consideration is a reason.[126] Here's Schroeder's central example. There will be a dance party tonight, and everyone is invited. Ronnie likes dancing, while Bradley hates it. That there will be dancing at the party looks like a reason for Ronnie to attend the party, but not one for Bradley (if anything, it's a reason for him to stay home).

Here, Schroeder is assuming – as is standard for internalists – that what explains Ronnie's reasons has to be a desire of Ronnie's. If others' desires are relevant, their relevance is explained in connection with some desire had by the agent who has the reason. For example, Bradley's desire not to dance does not on its own give rise to a reason for Ronnie not to attend the party,

124 Markovits (2014) p54-58

125 Schroeder (2007b) p203-208

126 Schroeder acknowledges that not all reasons are susceptible to this kind of contrast. There are some things we all have reason to do even if we don't have the relevant desires, and some things we we do not have reason to do despite having the relevant desires. We'll come back to these cases shortly. Nonetheless, Schroeder thinks, his methodological approach suggests that we should start with cases like Ronnie and Bradley's, then use these other cases as constraints on our theorizing. See Schroeder (2007b) p208-210.

though it could in light of some other desire of Ronnie's (like, for Bradley not to be stuck at home alone). In principle, though, this constraint could be relaxed and we would still have a recognizably internalist view. Kate Manne, for example, has argued that motivations like these for internalism could also be developed in favor of a view according to which my reasons are partially to be explained in terms of your desires and vice versa.[127]

In any case, Schroeder's methodological point isn't decisive on its own. There might be better ways of explaining why "there will be dancing at the party" has a different normative significance for Ronnie than it does for Bradley, and there might be cases where a difference in desires isn't the most natural explanation for a difference in reasons. Nonetheless, the contrast between Ronnie and Bradley offers a hint that internalism is on a plausible track.

## 2. Reductive Analyses and Methodological Background

One might respond to these suggestions with the worry that, whatever the moral or explanatory appeal of internalism, it just can't be a theory about justifying reasons because justifying reasons are normative and desires are not. These things might seem, as Enoch puts it, "just too different" for one to be reduced to the other.[128] This objection is a salient one for hybrid voluntarists too. Though they offer a different reduction of (some of) our reasons, if reductive accounts are non-starters, this looks like it will be as much a problem for reductions in terms of our desires as reductions in terms of our wills. So, while I'm going to go on to argue that there are problems for internalism, it's worth dwelling on how internalists can respond to this objection.

127 Manne (2016)
128 See Enoch (2011b) ch5 for one way of developing this idea.

This "just too different" objection might be developed in one of two ways. On one interpretation, it's a claim about whether *any* reductive thesis about reasons can be plausible, independently of the details of that purported reduction. Parfit presses a worry like this.[129] On his view, even though we might find out that there are two necessarily coextensive properties, one having to do with being a justifying reason and one having to do with desires, this would not amount to grounds for thinking that they were the same property, because the former is normative and the latter is not.

This might sound puzzling. Even when two properties have different senses (say, being heat and being molecular motion), it can turn out that one is reducible to the other. But Parfit responds that in the cases where this is possible, the property that gets reduced has a gap in it, waiting to be filled in by the property it is reduced to.[130] For instance, we might think that, before we knew about molecular motion, the property of heat could have been glossed as something like "the property, whatever it is, that causes solids to melt, gases to expand, and so on." If that's right, finding out that heat is molecular motion is just finding out how to fill the gap in our gloss on heat. Nonetheless, Parfit holds that normative properties are not gappy in this way. More specifically, he thinks that "being a reason to act" does not involve a "whatever it is" clause that could be filled in with some other property; being a reason to act is just normatively counting in favor of acting.[131]

I am suspicious about whether this version of the "just too different" objection gets off the ground. For one thing, as an epistemic matter, it's not clear to me why Parfit is so confident

129 Parfit (2011) and (2017).
130 Parfit (2011) p301-302.
131 Parfit (2017) p77

about which properties have gaps. As Mark van Roojen points out, Parfit's view seems to require that acquaintance with a concept is sufficient for knowing whether it's gappy.[132] But that seems implausible; I agree with Markovits that, without the benefit of hindsight, it's hard to tell which concepts are gappy.[133] Second, as a metaphysical matter, I think we should not find it implausible that properties with different senses could, in principle, turn out to be reducible to one another. The Evening Star and the Morning Star seem to be like this. Before discovering their identity, the senses of "Evening Star" and "Morning Star" didn't leave it completely open what instantiated them. They included the thought that different things instantiated them; we were just wrong about that.

So, I think, the idea that desires and reasons are just too different should not lead us to reject the possibility of any kind of reduction from one to the other. But the "just too different" objection can also be interpreted in a less flat-footed way. This second interpretation draws on the idea that some reductions seem to eliminate their objects, while some seem to explain them. To borrow an example from Markovits, if you discover that water is H2O, you should not think "Well, I guess there was never any water after all."[134] In contrast, if you discover that the monster under the bed is your brother, you *should* think "Well, I guess there was never a monster after all." Perhaps, then, the reason internalism strikes many externalists as a non-starter is that it's the kind of reduction that, if true, would eliminate its object.

What distinguishes these two kinds of reductions? As a general matter, it seems to have to do with how much the reduction asks us to change our views about the phenomenon being

132 van Roojen (2017)
133 Markovits (2014) p9
134 Markovits (2014) p10

reduced. For example, when we find out the heat is molecular motion, we need not change our mind about any of the familiar facts about heat, e.g. that it melts solids and expands gases, and that it hurts to touch hot things. But when you find out that the monster under your bed is your brother, you do have to change your mind about lots of your beliefs about the monster, e.g. that he will eat you, that he can't come out in the light, and that he has horns. Of course, a non-eliminative reduction needn't preserve *everything* we think about the phenomenon being reduced. We might learn, for example, that the Morning Star and Evening Star are not different entities. Still, we might think that a reduction is eliminative when it is very revisionary.

This diagnosis seems to be on the right track, but Schroeder notes that it can't be complete.[135] After all, we can always preserve the facts about a phenomenon by reinterpreting them. To borrow his example, consider a person who claims to believe in God, but goes on to say that what she means by that is that she believes in love. You might press this person by pointing out that while God and love have some things in common (say, they're both very powerful), there are lots of things that theists believe about God that are not true of love. Love did not create the universe, for example. This person might respond by saying that, on her view, love really did create the universe, and by "create the universe" she means "make the universe a worthwhile place to be." You might press again, coming up with a mismatch between her use of "create the universe" and the familiar one. But, in principle, it'll always be open to her to move the bump in the rug by redefining another term. Still, even though this person can make familiar claims about God come out true (at the expense of making them mean surprising things), it would be reasonable for a theist to treat her reduction as an eliminative rather than explanatory one.

135 Schroeder (2005)

Schroeder concludes that for a reduction to be non-eliminative, it can't just avoid being too revisionary about the phenomenon being reduced, it must also avoid being too revisionary about the phenomenon it is reduced to.[136] If we have a reduction like that, he proposes, it should not seem eliminative. So, the question facing the "just too different" objection is whether internalists can offer a reduction like that.

On the surface, it seems very clear that they can't. Intuitively, we can want things we have no reason to pursue. For example, after being cut off in traffic, you may want the offending driver to crash. But, having this desire does not make it the case that you have a reason to run the offending driver off the road. Following Schroeder, I'll call this the "Too Many Reasons" problem; it looks like internalism predicts reasons where there are none. We can also have good reasons to do things we don't want to do. For example, if you come across a drowning child, you have a reason to save her even if you do not want to do so. This is the "Too Few Reasons" problem; internalism seems to predicts that there won't be reasons where there are some. So, if internalism is to be an attractive thesis about the metaphysics of reasons, it will need to come paired with a strategy for addressing these extensional problems. In this way, there is extra pressure on internalists to accommodate intuitions about reasons rather than to argue for reforming them. Too much reform risks changing the subject by making the internalists' reductive project an eliminative one. As a result, a dispute between an externalist and an internalist about the metaphysics of reasons differs from a dispute between two externalists with different normative theories about whether to accommodate or reform a particular intuition. Because of their metaphysical ambitions, internalists have to be more cautious about reform.

136 Schroeder (2005)

It's not surprising then, that getting the extension of reasons right has been a central task for defenders of internalism. In what follows, I'll canvass some of the most ambitious and intriguing attempts to do it. I'll argue that while these strategies make progress, ultimately some version of the familiar extensional worries remain for these views. As a result, internalism doesn't pass the test for a non-eliminative reductive theory. But, its failure to pass is contingent on its not getting the extension of reasons right. If a reductive theory (or partially reductive hybrid theory) did get the extension of reasons right, it wouldn't be objectionable merely for being a reductive (or partially reductive) theory. So, the bad news for internalism is good news for hybrid voluntarism.

My argument against internalism, then, will hinge on the Too Many and Too Few reasons problems, i.e. on the claims that the extension of the theory is both over-inclusive and under-inclusive. This argument therefore depends on claims about what reasons we have, claims about which I'm more confident than I am about any particular metaphysical theory about in virtue of what we have those reasons. Still, I hope that the objections I develop here can be at least somewhat ecumenical across different first-order theories about what reasons we have. For example, while many philosophers who press the Too Few reasons problem are chiefly concerned with explaining *moral* reasons (e.g. that we have a reason to help those in need even when we don't want to), the Too Few reasons problem doesn't necessarily have to rely on intuitions about reasons to be moral. It just has to rely on intuitions that there are some reasons we have even when they are not supported by desires. These reasons could instead be prudential (e.g. reasons to go to the gym even when we don't want to) or aesthetic (e.g. reasons to learn art

history even when we don't want to) or whatever. So, while I'm going to rely on intuitions about what we have reason to do, I don't think I need to rely on any *particular* intuition of that sort. Depending on your views about what we have reason to do, different sorts of examples could fit the bill.[137]

Another caveat: because of the tendency to think of the Too Few reasons problem in terms of moral reasons, it might seem that the Too Few reasons problem is a more important one than the Too Many reasons problem. If internalists could show that we have reason to be moral even when we don't want to, perhaps it wouldn't be so bad if they had to allow that we have reasons to do other things it initially seems like we don't. I am skeptical about this. It seems to me that the categories of reasons that can motivate the Too Few reasons problem can motivate the Too Many reasons problem just as well. For example, getting the extension of moral reasons right isn't just a matter of explaining why we have reason to help others; it's also a matter of explaining why we don't typically have reasons to hurt them. Similarly, getting the extension of prudential reasons right isn't just a matter of explaining why we have reasons to go to the gym; it's also a matter of explaining why we don't have reasons to go on an all-cake diet. So, it seems to me, concern to account for a particular domain of reasons shouldn't lead us to put any extra theoretical weight on either kind of extensional problem.

## 3. Varieties of Internalism

What might internalists say to avoid these problems? One kind of strategy for securing

137 Here's another way to put this point. Extensional objections to internalism are often made by appeal to moral rationalism, but you don't have to be a *moral* rationalist to accept them. As long as you're a rationalist about some class of reasons, then those reasons will be hard for internalists to accommodate.

extensional adequacy involves *idealization*. Rather than explaining reasons in terms of a particular agent's desires, we might try to explain reasons in terms of that agent's desires after they have been cleaned up a bit. The *locus classicus* for this kind of strategy is Bernard Williams' paper "Internal and External Reasons." There, he notes that a version of internalism without any idealization would face two kinds of problems. First, desires can be based on false beliefs, and these desires do not seem to give rise to reasons. I may want to drink the contents of a glass because I believe those contents to be gin, when in fact the glass contains petrol.[138] Second, desires can be related to actions at some remove. If I want to visit Colorado, I may have a reason to buy a plane ticket, even if I do not want specifically to buy a plane ticket.

According to a common reading of Williams,[139] his considered view is that internalists should accept two forms of idealization. First, they should exclude desires based on false beliefs. This is intended to solve the gin/petrol problem. Second, they should include desires that an agent would have after following a "sound deliberative route" from their current motivational set, and perhaps exclude desires that they would lose after following such a route.[140] This is intended to account for the Too Many and Too Few reasons problems. Following a sound deliberative route might produce desires to do things like take the known necessary means to our ends, and might rid us of some desires, like those that turn out on reflection to be hard to square with others that are more important to us.

---

138 This is a classic example in metaethics. But, as Carlos Pereira Di Salvo pointed out to me, it's totally implausible. The smell of petrol would be a dead giveaway. So, feel free to amend it as needed to make it plausible; maybe it's an odorless-but-still-bad-to-drink liquid, or maybe I lack a sense of smell, or whatever.

139 Finlay (2009) raises some exegetical questions about how we should interpret Williams' view, and even whether he's addressing the question that subsequent Williams-inspired internalists have been. Since my primary concern here is the space of possible internalist views rather than where Williams himself fits in that space, I'm going to focus on the common reading.

140 Williams (1979) p105

There is a lot about this proposal that could be filled in more fine-grainedly. For one thing, it's hard to say exactly what a sound deliberative route is supposed to involve. It's clearly meant to include things like instrumental reasoning, but one might wonder how much else is involved, and how much revision to the original set can result. It's also not totally clear what an agent's motivational set includes in the first place. Desires are in, but what about other pro-attitudes, goals and personal policies, or normative beliefs? Finally, it's hard to say exactly how accommodating Williams wants to be when it comes to intuitions about reasons; he uses his internalism to cast doubt on the idea that we really do have reason to do some of the things we tend to think we do.[141]

Thankfully, subsequent internalists have worked to show us how strategies like the ones Williams suggests might help to address internalism's extensional challenges. So, rather than trying to work out what Williams himself would say, I'll move on to considering their specific proposals. One such proposal comes from Peter Railton. On Railton's view, our reasons are a function not of our own desires, but by a relation between our desires and an idealized set.[142]

Railton formulates this view in terms of what's good for an agent, rather than that agent's reasons. So, strictly speaking, the way I'm about to recast it might not perfectly preserve Railton's thinking. In particular, if what we have reason to do outstrips what's good for us, then we can't move directly from a view about what's good for an agent to a view about what that agent has reasons to do. Nonetheless, Railton's motivation is to see how far an internalist can go

141 For example, in Williams (1989), he suggests that a husband who is not motivated – and couldn't be motivated by following a sound deliberative route – to be kind to his wife doesn't in fact have any reasons to be kind. This is a terrifying result; if Williams were really trying to arrive at an extensionally adequate version of internalism, this seems like exactly the kind of case where he would want to be able to accommodate our ordinary judgment.

142 Railton (1986)

in accommodating our typical judgments about reasons, so I think my way of presenting things here preserves the spirit of that project.

Anyway, here's how Railton's idealization works. Imagine a better-informed version of me, Cory+. Cory+ knows all the non-normative facts, including being vividly aware of phenomenological facts about what things will be like for me to experience, and he meets whatever formal constraints on rationality you like (e.g. his ends are coherent and mutually supportive).[143] Presumably, Cory+'s desires will be quite different from my own, given that I'm not fully informed or perfectly structurally rational. On Railton's view, my reasons are to be explained in terms of what Cory+ would want me to want. Let's call such desires "ideally endorsed" desires. We can then say that, for Railton, I have a reason to do something just in case doing that thing would promote the satisfaction of one of my ideally endorsed desires.

The distance between me and Cory+ opens up space for replies to both the Too Many and Too Few reasons problems. Take the Too Many Reasons problem first. Sometimes I find myself idly wanting to hike the Pacific Crest Trail. I'm not a skilled hiker, nor do I have the wilderness survival skills required for such a project. Nonetheless, in certain moods hiking the PCT sounds fun to me. If we take a flat-footed version of internalism, it looks like I have a reason to hike the PCT, given that doing so would promote the satisfaction of this desire. But this seems like the wrong result; I don't really have a reason to hike the PCT. Railton can respond that my desire to hike the PCT isn't one that Cory+ would want me to have. Plausibly, learning more about the

143 The restriction to "non-normative" facts is meant to preserve the reductive credentials of this view. Cory+ doesn't know facts about what reasons I have, since those are the facts Railton will be using him to explain. The restriction to "formal" constraints on rationality is meant to rule out substantive views where being rational would involve taking into account the normative facts. For instance, if assuming that Cory+ is rational required assuming that he desires only what he has reasons to desire, we'd again sacrifice the reductive potential of this view.

difficulty of the task and the skills required would disabuse me of this desire. So, it's not in the ideally endorsed set, and so not the kind of desire that can explain reasons on Railton's internalism.

Now consider the Too Few Reasons problem. It often happens that I don't want to go for a run, so a flat-footed version of internalism would need to say that I don't have a reason to do so. Nonetheless, it seems plausible that I do. Railton's version of internalism, though, offers two resources for addressing objections like these. First, Cory+ is ideally rational, so for any end of mine that's ideally endorsed, he'll also want me to want to take the means to it. So, given that I do want some things I can get from going for a run (e.g. cardiovascular health, mental clarity), he'll want me to want to go for a run. Second, Cory+ is vividly aware of what all possible experiences will be like for me. Were I vividly aware of what going for a run would be like (say, the improved mood later in the day) I might want to go. In both of these ways, the set of ideally endorsed desires can include desires that we don't currently have, and those desires are candidates to explain reasons to do things we don't currently want to do.

So, Railton's view offers internalists new resources for addressing the Too Few and Too Many Reasons problems. It's not clear, however, that these resources are sufficient. For one thing, each of these strategies relies on some guesses about what Cory+'s desires are like. We expect that he wouldn't want me to want to hike the PCT, and that he would want me to want to go for a run. But Cory+ is *very* idealized, so we may not be in a position to guess how things will turn out on the other side of idealization. Perhaps, once one is vividly aware of what everything will be like, nothing seems worthwhile anymore, and Cory+ wouldn't want me to want anything.

Or perhaps I have a weird disposition, such that knowing more about the dangers of the PCT would only increase my desire to hike it.[144] We could rule these possibilities out in advance if we built something about reasons (say, to develop my capacities or to avoid danger) in to the idealization procedure we used to define Cory+, but once we do that we'll no longer be offering a purely internalist view about my reasons. As it stands, Railton's strategy may address internalism's extensional challenges, but whether it does rests on a big empirical bet.

And even if the empirical bet works out, this approach has some deeper problems. Connie Rosati identifies two. First, it's not clear whether the kind of idealization Railton has in mind is conceptually possible. A big part of Cory+'s advantage over me is that he knows what certain experiences are like in ways that I can only imaginatively simulate. But having this advantage for some experiences seems to preclude having it for others.[145] For example, like many philosophers I am disposed to process new information in a particular way: to look for arguments, to evaluate those arguments in abstraction from who is offering them, to figure out where this idea fits in conceptual space, etc. These, of course, are not the only ways of reacting to new information. One might also look for relevant personal experiences to share, or identify the motives of the person offering the information, or look for ways to make the person speaking feel heard. These aren't strictly incompatible, but they are different defaults; doing one when you're in the habit of doing the other requires effort and attention, and that changes what it feels like to do.[146] Cases like this suggest that Cory+ can't really be vividly informed about what *everything* will be like

144 Arneson (1999) makes this point, drawing on Gibbard (1990) p20.

145 Rosati (1995a)

146 Of course, this is an oversimplification for the sake of having a clear example. Any real person's defaults are more complicated. For instance, how and when the philosophical approach kicks in depends on things like who I'm with and what we're talking about. Nonetheless, the point that we have these dispositions and that they're relevant to what different experiences feel like for us will still be true even when applied to more realistic cases.

for me; if he's vividly aware of what it's like to be in some situation as a philosopher, he can't also be vividly aware of what it's like to be in that same situation with some other set of defaults (perhaps we could imagine that he lives a life with one kind of psychology and then another and then compares them, but the problem recurs when we ask what his psychology is like when he makes the comparison).

Second, the same kinds of cases suggest that Railton's idealization strategy is too local to give a general answer to the Too Many and Too Few Reasons problems. Some decisions are about what kind of people to be, rather than what to do given the kinds of people we are.[147] Idealizing from the desires we have now, then, might stack the deck against change inappropriately. For instance, many people cite having children as the kind of experience that changes their preferences and sense of who they are. So finding out what Cory+ thinks about the possibility of my having children is only so helpful; I may also want to know what Parent+ thinks, where Parent+ is the result of idealization from the desires I would have after becoming a parent.[148]

Another worry concerns whether this view can explain the modal strength of judgments about reasons. Take the fact that going for a run will improve my cardiovascular health. It might seem that this counts in favor of going for a run in a very modally robust way, perhaps necessarily, or in all possible worlds in which I'm a creature who depends on a cardiovascular

147 Rosati (1995b)

148 It's not clear that one's future preferences are *always* the right ones to take into account. Some ways that preferences change might lead us to doubt their normative significance. For instance, if I know that after being kidnapped I will have Stockholm syndrome, finding out that in the future I will prefer to be held captive should not make me think it'll be good for me to be held captive. I'm not sure what the right general story is about when we should discount future preferences and when we shouldn't, but all I need here is the minimal point that sometimes we should take them seriously.

system. But it might seem much more contingent that Cory+ would want me to want to go for a run, particularly after we reflect on how weird Cory+ has to be after all the relevant idealizations are done. So, even if Railton's view gets individual cases right, it might not get them right in nearby possible worlds.[149]

While Williams and Railton try to address internalism's extensional challenges by idealizing from the desires agents already have, others have tried to get the extension of reasons right working only with agents as we actually find them. One version of this strategy comes from Velleman. On his view, part of being an agent is having a particular desire. Specifically, it's a desire to understand ourselves, where this means being able to explain what we're doing in folk-psychological terms, i.e. in terms of our beliefs, desires, habits, ideals, personality traits, and the like.

Velleman's arguments for this surprising claim come mostly from action theory; he thinks that the best way out of various puzzles about how to distinguish actions from mere happenings, how intentions can guide agents over time, and how to explain the normativity of various demands of rationality is to accept that we have this desire just in virtue of being agents[150] (even if it's hard to find it introspectively).[151] Nonetheless, if this story about action is right, it promises to help internalists with the objections we considered above.

Velleman's story of rational action starts with the observation that, given a desire for this kind of self-understanding, there are two ways to go about satisfying it. One is to conform our narratives about ourselves to what happens: we might see patterns in our behavior and come to

149 Thanks to David Brink for suggesting this objection.
150 See Velleman (2000), Chapters 6, 7, and 10.
151 To be honest, I find this claim of Velleman's very intuitively plausible, but my sense is that many philosophers have the opposite reaction, so I'm just taking their word for it here.

form a self-conception in light of them. For example, I might notice that I spend a lot of time thinking about what would be fun to cook, and come to think of myself as an adventurous cook. Another strategy is to act in ways that flow from our existing self-conception; this way, we can see our resulting behavior as expressive of who we think we are.[152] For example, having come to think of myself as an adventurous cook, I might decide to spend my afternoon learning to make a new dish rather than listening to new music.

This second way of attaining self-knowledge has the consequence that, whenever a course of action will conform with my self-conception, pursuing that course will contribute to the satisfaction of my desire for self-understanding. When I'm evaluating some possible options, I can see them as expressing the total of my self-conception more or less well. And given my desire to understand myself, the features that make them do this better come to be reasons to act (in that acting that way would promote self-understanding) while the things that make them do this worse count as reasons not to (in that acting that way would make it harder to explain myself).[153] For example, maybe giving someone I haven't seen in a long time a bear-hug expresses my excitement to see them, but not my uncertainty about how close we still are, while other possible greetings navigate between these more coherently. That I'm excited counts as a reason for the hug; that we haven't been in touch counts as a reason against it.

If this story is correct, it helps with both of the problems for a flat-footed version of internalism. Consider the Too Few Reasons problem. While I might not have a desire to do go for a run, it's likely that some part of my self-conception counts in favor of it. Maybe I think of

152 See Velleman (2009) p17-18
153 See Velleman (2009) p18

myself as prudent or not lazy, and failing to go for a run would be hard to square with these things.

Velleman can address the Too Many Reasons problem in a similar way. When I find myself idly wanting to hike the PCT, this desire may give me a reason to do so. But this desire is a tough one to square with the rest of my self-conception, given the rest of the facts about me and my unpreparedness for that particular challenge. So, acting on that desire would tend to frustrate the goal of acting in ways I can understand.[154]

Of course, all this only works if you accept Velleman's claim that all agents have a desire for self-understanding. If one could be an agent without having this desire, then Velleman would fall prey to the same kinds of problems I raised for Korsgaard in Chapter 2; if there really could be a Perfectly Consistent Caligula, for example, who only desired harming others, then Velleman's arguments from the desire for self-understanding to a solution to the Too Few Reasons problem wouldn't get off the ground, since Caligula – by hypothesis – doesn't have the desire for self-understanding.

For my purposes here, we can safely set these questions about Velleman's constitutivism aside. That's because, I think, even if it's true that agency constitutively requires a desire for self-understanding, this is insufficient for solving internalism's extensional problems. The difficult cases for Velleman are ones where an agent is irrational, but in familiar enough ways to be self-aware about it. Consider an agent who, based on observation of her past behavior, takes herself

154 I've illustrated Velleman's responses to the Too Few and Too Many reasons problems with non-moral cases here because those are the ones that Velleman himself is most concerned to address. He is ambivalent about how far his view can go in accommodating moral reasons. See Velleman (2009) p149-151, and for a more fully developed version of this view, Velleman (2013). Nonetheless, as I'll explain below, I think we can see issues with his view even if we stick to the kinds of cases he's focused on.

to be lazy. As a result, lazy actions (like staying home and watching TV) are easy to square with her self-conception. If so, it looks like Velleman has to say that her desire for self-understanding counts in favor of doing them. But, plausibly, that something would be lazy doesn't justify doing it. So, Velleman seems stuck with the Too Many Reasons problem.

Velleman anticipates this worry.[155] His response is that self-ascriptions of negative character traits do not really contribute to the intelligibility of acting in conformity with those traits.[156] His explanation for this is that such self-ascriptions involve a judgment that the actions they support are not really worth doing. For instance, what makes an action lazy by an agent's own lights is that it would be better not to do the action. If our agent saw value in staying home and watching TV, she'd see it as relaxation or self-care rather than laziness. So, when an agent attempts to make sense of herself by self-ascribing a negative character trait, there will be a tension in her self-conception. She'll at once see some action as worth doing (because it expresses the trait she thinks she has) and not worth doing (because of the negative judgment involved in that trait).[157] If so, acting this way doesn't help us make sense of ourselves after all.

I don't think this reply works. Velleman might be right that some negative self-ascriptions work this way, but I'm not sure that all of them involve outright inconsistency. Suppose that an agent's laziness works in the following way. When she's reflecting on a specific question, like what to do this Saturday, the options that come to mind are lazy ones: watching TV, napping, etc. But, when she's reflecting on her life in a more abstract way, she sees other things she spend her time on, and wishes that her laziness didn't screen off these possibilities during her in the

155 Velleman (2009) p31-33
156 Velleman (2009) p31-33
157 Velleman (2009) p31-33

moment deliberations. Such an agent might, on the one hand, think of staying home as her best option given the set of things that make it on to her deliberative stage, and, on the other hand, think of staying home as not her best option given the full set of things that she could do. This is a sad situation to be in, but not an inconsistent one. So, an agent whose laziness operates by selecting which options make it to the heat of deliberation, and who knows this about herself, looks like a counterexample to Velleman.[158]

The source of the problem for Velleman here is that consistency comes cheap. When we recognize some tension between parts of our self-conception, or our self-conception and some bit of action, we can resolve the tension either by bringing the two into line, or by making a distinction (like between a local perspective on some bit of deliberation and a global one). And while making the distinction might come with some costs to intelligibility, those costs might be outweighed by ease of fit with our observations about our behavior. But as the lazy agent shows, a self-conception that fits with our behavior might not be one that justifies it. The upshot of these considerations is that while a desire for intelligibility puts some pressure on us to act in recognizably rational ways, it's too weak a requirement to solve the Too Many reasons problem. There are ways of achieving self-understanding that still run afoul of our ordinary judgments about agents' reasons.

Structurally similar problems arise for Markovits' version of internalism. Markovits sees the main challenge for internalists as a moral version of the Too Few Reasons problem; Scrooge has reasons to give to charity, even if he doesn't want to.[159] Her strategy for addressing this

158 I'm using laziness here because it's Velleman's example, but it might be that the kind of counterexample described here is easier to imagine with other kinds of negative self-ascription. For example, boredom can be like this. When you're bored, it's harder to think of the non-boring things to do you normally recognize.

159 Markovits (2014), ch6

challenge draws on familiar Kantian resources, though ones not often associated with internalism. In particular, she suggests that an agent's desires can be susceptible to coherence pressure in much the same way that beliefs can. So, just as we might say that if I believe my own well-being is valuable and believe that you are an agent just like me, there's coherence pressure to believe that your well-being is valuable as well, Markovits wants to say that if I desire my own well-being and believe that you are an agent just like me, then there's coherence pressure to desire your well-being as well.[160]

Markovits may well be right that desires are susceptible to coherence pressure like this. But, just like for Velleman, coherence pressure only goes so far. One might achieve coherence in one's desires by coming to desire others' well-being, or one might achieve coherence by introducing a distinction that makes one's own well-being special. This latter move might be unprincipled or brutish, but it's not incoherent. So, it's not clear what Markovits can say about why an agent who goes this way still has reasons to be moral.

Schroeder offers another approach to these issues. The core of his view takes the same form as the flat-footed version of internalism with which we started: you have a reason to do something whenever doing that thing would promote the satisfaction of one of your desires, and the content of that reason is a consideration that picks out the contribution that action would make toward satisfying that desire. For example, the consideration “there is coffee in the kitchen” is a reason for me to go to the kitchen in virtue of the fact that I want a coffee.

This way of formulating internalism looks like it should immediately give rise to the Too Few and Too Many Reasons problems, and without the resources for mitigating them proposed

160 Markovits (2014), ch6.

by Railton, Velleman, and Markovits. Nonetheless, Schroeder thinks these problems can be avoided if we're careful about how we understand the idea of "promoting" the satisfaction of a desire, and in how we spell out the way reasons combine to result in judgments about what we ought to do. Spelling out his view requires a fair bit of formal machinery, so I'll first try to get all the pieces on the table and then assess how far they go toward getting the extension of reasons right.

The first step in Schroeder's response to the Too Few and Too Many Reasons problems is to refine our judgments about when there are reasons and when there are not. He points out that when we talk about reasons, there are pragmatic norms that direct us to only mention reasons that are weighty enough to make a difference in our conversational context.[161] For example, suppose you are choosing between three routes you might drive home, and it's important to you to make it home before the World Series starts. The first route is ugly but will get you home on time, the second is more scenic but comes with a substantial risk of being late, and the third is equally scenic but will all but assure that you are late. Given these options, it'd be natural for me to say "You have no reason to take the second route." What you care about is getting home on time, and the first route dominates the others when it comes to that. But, it's not literally true that you have no reason to take the second route. After all, you have more reason to take the second one than the third.[162]

What's going on here, Schroeder hypothesizes, is that the consideration that counts in favor of the second route, that it offers you a decent chance of making it home on time, doesn't

161 Schroeder (2007b), p92-97
162 Schroeder (2007b), p94

make a difference to whether you ought to take the first route when choosing among the three. So, it's natural for us to speak loosely and say that there's no reason to take it. But, when we change the context to raise the standards of precision (i.e. when we ask, "so there's really *nothing* the second route has going for it that the third route doesn't?"), we drop the loose speech and note the reasons that count in favor of the second route but not the third.[163]

Schroeder notes that when we drop the expectation that we will only mention reasons that could make a difference in our deliberative context, it turns out that reasons come very cheaply. For just about any action, it's possible to say something about how that action would promote the satisfaction of one of my desires, even if would do so very indirectly or weakly.[164] For example, just about nothing counts in favor of attempting to eat my car, but at least this much does: it will satisfy my daily recommended dose of iron.[165] This sounds weird, but it sounds weird in precisely the way Schroeder's hypothesis suggests it should.[166] Given that this reason could not possibly make a difference to what I decide about whether to eat my car, it's the kind of thing we'd typically ignore when talking about my reasons. In a different conversational context, say one where we're comparing eating my car to eating my shirt as part of an improv game, weak reasons like this aren't pragmatically screened off.

These two points together suggest an alternative strategy for internalists. Rather than trying to show that we have, strictly speaking, no reason to satisfy some of our desires (like, my desire to hike the PCT), and some reasons to do things we don't desire (like Scrooge's reasons to give to charity), internalists could instead say that we have a reason to do pretty much anything.

163 Schroeder (2007b) p92-97
164 Schroeder (2007b) p112-113
165 Schroeder (2007b) p96
166 Schroeder (2007b) p96

The challenge, then, is to construct an account where reasons of the former sort are typically not weighty enough to mention, but reasons of the latter sort are. This would explain our intuitions about these cases in a way that's compatible with the idea that reasons, in an unrestricted sense, come cheap. So, for Schroeder, the central challenge for internalists is coming up with a way to spell out the weight of reasons that can accommodate these two ideas.[167]

Recasting questions about the existence of reasons in terms of the weight of reasons may not seem like much progress. After all, my reason to hike the PCT and Ronnie's reason to go to the dance party are both relatively directly connected to our desires, but intuitively Ronnie's is weighty and mine is not. Similarly, Scrooge's reason to give to charity and my reason to eat my car are both relatively indirectly connected to our desires, but intuitively the former is weighty and the latter is not. So, to address the Too Many and Too Few Reasons problems, an account of the weight of reasons is going to have to provide resources for distinguishing these pairs of cases.

It's tempting to think that for internalists, the weight of a reason has to correspond to the strength of the desire that explains it, or to how good a means the action is to the desired end. Schroeder argues that, if we want any hope of overcoming the Too Many and Too Few reasons challenges, we need to resist this temptation. His alternative starts with the thought that when we

167 One might worry that this strategy undercuts some of the initial motivations for internalism. Admitting that strictly speaking reasons come cheap might, for example, make it harder to see a tight connection between normative beliefs and motivation to act. For philosophers who are attracted to internalism on these grounds, I think it's right that Schroeder's approach should be unsatisfying. But, I don't think – nor do I think Schroeder thinks – that those are the best ways to argue for internalism. The other motivations I mentioned at the start of this chapter, e.g. worries about the metaphysical commitments of externalism, the hope of offering a reductive alternative, and considerations about what it's like to reason with another person are consistent with Schroeder's strategy.

ask how weighty a reason is, that question can be reinterpreted as a question about other reasons.[168] For example, if I want to know how much weight to put on the consideration "eating my car will fulfill my daily dose of iron," what I need to do is think about what counts in favor or against putting weight on it. I might note, for example, that there are lots of other ways of getting my daily dose of iron. So, Schroeder proposes, we can answer the question about how weighty a reason is in terms of the existence of other reasons to place weight on it or not.[169] In other words, on Schroeder's view, asking whether a reason is weighty is like asking whether a person is admirable. What makes a person admirable is the reasons there are to admire her, not the strength of the admiration anyone in particular happens to feel.

This observation helps Schroeder in two ways. First, it preserves the reductive credentials of his internalism. If the weight of reasons can be explained in terms of the existence of reasons, and the existence of reasons can be explained in terms of desires, then we still have an account that at bottom directs us toward desires. No extra resources, like brute facts about how weighty some reasons are, need to be wheeled in. Second, this understanding of weight rules out a tempting alternative, where the weight of a reason just depends on the strength of the desire that explains it. Instead, on Schroeder's view, desires only explain the existence of reasons; the weight of a reason is always function of what other reasons there are. As a result, an agent's reasons depend for their existence on her desires, but the balance of reasons doesn't depend on what she wants most. This is a good consequence. Without it, we could redeploy the Too Many Reasons problem by imagining cases where an agent wants something badly, rather than just

168 Schroeder (2007b) p129
169 Schroeder (2007b) p132

wants it.

Still, a lot hangs on the details of how a reason's weight is determined. The first thing to notice is that, even though the weight of a reason depends on further reasons to put weight on it, not just any kind of reason will do. For example, if you offer me a million dollars to place weight on the reason "eating my car will get me my daily dose of iron," that doesn't make it any weightier of a reason. Intuitively, your offer is a reason of the *wrong kind*. So, to a construct an account of the weight of reasons, Schroeder first has to screen off reasons like these.

Schroeder notes that this challenge is structurally similar to ones that show up in other areas of philosophy where an activity seems to presuppose norms about what is and isn't a good reason for engaging in that activity. For example, in Pascal's Wager, the benefits of an afterlife seem like the wrong kind of reasons to believe in God (belief is supposed to be sensitive to the truth), and in the toxin puzzle, the financial benefits of intending to drink the toxin seem like the wrong kind of reasons to intend to drink it (intention is supposed to be sensitive to the reasons for doing the intended action).[170] Schroeder proposes that in general, what we should say about these cases is that reasons are of the right kind when they are reasons that anyone would have, just by virtue of being involved in the relevant activity.[171] For example, the right kind of reasons to believe are reasons that anyone would have to believe, just in virtue of being an epistemic agent (and not, say, the reasons she has in virtue of her practical interests). Similarly, if I'm shooting pool with a friend who is just learning the game, some of my reasons (say, not to sink my opponent's balls) are ones that anyone would have insofar as they're playing pool, while

170 Schroeder (2007b) p135-136
171 Schroeder (2007b) p135-136

some of my reasons (say, to keep the game relatively close) are ones that one could be a pool player without recognizing (e.g. one who is playing in a tournament rather than with a friend).

So, Schroeder concludes, the reasons we should take into account when assessing whether some reason is weighty are only those that any practical agent would take into account, just in virtue of being a practical agent. How do these reasons combine to produce a judgment about weight? It'll help to start with an example, and generalize from there.

Suppose I'm deliberating about whether to give a friend a ride to the airport, and am considering how much weight to place on the fact that airport parking is expensive. I then remember that she has a parking permit for an airport hotel that will allow her to park there for free. This looks like a reason not to place weight on the fact the parking is expensive. But I might then remember that her parking permit has expired. This looks like a reason not to place weight on the fact that she has the permit, and so not to fail to place weight on the fact that parking is expensive. I might then remember that the hotel often doesn't check the expiration date on parking permits. This looks like a reason not to place weight on the expiration, and so to place weight on the existence of the permit, and so not to place weight on the expense of parking at the airport.

This chain of undercutting defeaters could go on for an arbitrarily long time (maybe I find out that the hotel has a new parking enforcement scheme...) but it can't go on forever. Eventually, one of my reasons for placing weight on some other reason will not be undercut. And once that happens, all we need to do to answer my original question about whether to put weight on the fact that airport parking is expensive is to zip up the chain of undercutting defeaters. Sometimes

the chain will be long, and sometimes it won't (perhaps there's no special permit to begin with), but in either case whether I ought to place weight on the expense of airport parking depends on whether or not any reasons ultimately undercut doing so.

We can put this point more generally if we apply a recursive rule. If we want to know whether to place weight on some reason R, we need to look for further reasons S, T, etc. not to place weight on it. If there are none, R is weighty. If there are some, we need to ask whether they are weighty, until there are no further reasons to consider. Then we zip up the recursive chain and arrive at an answer about the weightiness of R.[172]

Putting all these ideas together: Schroeder's account says than a reason is weighty when it is not undercut by any reasons of the right kind. We can now apply this view to the extensional problems for internalism and see whether Schroeder has made progress.

Take the Too Many Reasons problem first. There, the challenge was to distinguish between Ronnie's reason to go to the dance party (which looks weighty enough to mention) from my reason to hike the PCT (which doesn't). The recursive strategy looks like it should work here. Absent further details, there's no reason for Ronnie not to place weight on the fact that there will be dancing at the party. So, it looks like a perfectly good reason for him to attend. But, there are plenty of reasons not to place weight on the fact that hiking the PCT sounds fun; I don't have the requisite skills to do it nor the experience necessary to evaluate whether it really would be fun. So far so good for Schroeder.

Now consider the Too Few Reasons problem. There, the challenge was to distinguish between Scrooge's reason to give to charity (weighty) and my reason to eat my car (not weighty).

172 Schroeder (2007b) p138

The recursive story does a good job with my reason to eat my car; there are plenty of reasons not to place weight on that. Things are trickier with Scrooge's reason to give to charity. After all, Scrooge wants to keep his money, and that looks like a *prima facie* reason for him not to place weight on facts like "this money could provide Tiny Tim a Christmas goose." But, Schroeder argues, reasoning like that would put Scrooge in violation of the right kind of reasons constraint.[173] That constraint requires that we approach questions about the weight of reasons in a way that any practical agent could approach them. But Scrooge's preferences are idiosyncratic here; not everyone is so frugal and not everyone is so indifferent to others' lack of a Christmas goose. So, Scrooge's desire to keep his money isn't an undercutting defeater after all.

At least, that's how Schroeder's account of weighing reasons is intended to go. If it works, it provides an elegant solution to the Too Many and Too Few Reasons problems. Unfortunately for the prospects of internalism, though, I don't think it works. I see two problems.

The first problem concerns the right kind of reasons constraint. The function of this constraint is to make sure that agent-neutral reasons (like Scrooge's reason to give to charity) can be weighty even when they run counter to what an agent most wants. The constraint fulfills that function by ensuring that agent-neutral reasons can't be undercut by agent-relative defeaters, because – by virtue of being agent-relative – those defeaters won't be the kinds of reasons any agent would take into account. This gets us the right result for cases like Scrooge, but I'm not sure it's the right result more generally.[174] Consider Bradley and the dance party. Recall that Bradley doesn't like dancing, and so "there will be dancing at the party" is not a reason for him to

173 Schroeder (2007b) p142
174 Thanks to Matt Braich for discussion on this point.

go in the way it is for Ronnie; in fact, given Bradley's apprehensions about dancing, it's a reason not to go. Imagine that, nonetheless, there is some agent-neutral reason for Bradley to go to the party. Perhaps there will be some people there it would be prudent for him to get to know. Structurally, this case is a lot like Scrooge's: there is some agent-neutral reason to act, and an agent-relative reason that might serve as an undercutting defeater. But the two cases seem very intuitively different. That Scrooge doesn't want to give up his money doesn't undercut his moral obligations, but that Bradley doesn't like dance parties *does* seem to undercut his prudential reason to go to the party. "I don't like dancing" is a perfectly fine justification for not putting much weight on the professional benefits of attending the party.[175]

The upshot of this comparison is that the relationship between agent-neutral reasons and agent-relative undercutters is more complicated than Schroeder's right kinds of reason constraint allows. Sometimes agent-relative reasons are good undercutters and sometimes they're not. I'm not sure what to say in general about when they are; some prudential reasons to attend the party really do seem not to be undercut by Bradley's preferences (say, if Bradley is offered lots of money to attend) and some moral ones do seem to be undercut (say, if attending would merely brighten Ronnie's day). But whatever the right relationship between these is, it's not captured by Schroeder's right kinds of reasons constraint. Sometimes it is appropriate to approach questions about how much weight a reason has from the perspective of your own idiosyncratic preferences. So, there's a dilemma for Schroeder: include the constraint and get cases like Bradley's wrong, or

175 If you're not sure about this, imagine that you're the graduate representative advising the department on how to create an inclusive atmosphere. You will suggest that it's important that there be opportunities to network that don't take place at dance parties, because doing otherwise would be unfair to people like Bradley. If it makes the case easier to imagine, you can substitute parties with dancing for parties with drinking, or parties where one can't bring one's young child. If preferences like Bradley's were not good reasons for not attending such parties, it wouldn't be unfair to concentrate professional benefits at them.

leave it out and get cases like Scrooge's wrong.

The root of this problem, it seems to me, is the argument Schroeder gives for introducing the right kinds of reasons constraint in the first place. Recall that his idea was that in other contexts, the right kinds of reasons are the reasons one would have just insofar as one is involved in a specified activity, rather than reasons one might have insofar as one is involved in that activity *and* has other concerns. That's why prudential reasons are the wrong kind to settle whether to believe in God; your prudential concerns aren't ones you have just in virtue of thinking about theology. Nonetheless, it's not clear to me that deciding what to do is something agents do just in virtue of the fact that they are agents. Instead, it's something we do in virtue of being agents with particular ends. Without specifying any ends, it's hard to say a lot about what agents will take into account.[176] So, even if we think Schroeder has the right schema for answering questions about what the right kinds of reasons are in some domain, it's not clear that the schema counts in favor of the constraint he wants to apply to weighing reasons.

Of course, it could be that there's some other way of spelling out the right kinds of reasons for assessing weightiness that does not encounter these problems. I can't think of one, but I haven't ruled it out. So, it'll be instructive to give Schroeder the benefit of the doubt here. What happens if we grant that the recursive story about weight can distinguish Scrooge and Bradley?

Unfortunately, I think there's a second problem for Schroeder's account of weight. The problem is that not all of our questions about how to weigh reasons are questions about whether a particular reason is weighty; we also sometimes need to know, of two weighty reasons, which

176 In saying this, I don't mean to rule out the possibility that there are some things we can say about what an agent should decide absent any information about her particular ends. Maybe Kant (1785) is right that some ends are obligatory for rational agents. Still, when we decide things, we don't just decide as bare agents with whatever obligatory ends that requires. We decide as agents with those ends plus the ones we have set for ourselves.

is weightier. To keep things simple, consider a case with no undercutting defeaters, and only one reason on each side. For Ronnie, that there will be dancing at the party is a reason to attend. But it might also be that Ronnie has a reason not to attend; maybe he has a paper to write. If neither of these reasons is undercut, following Schroeder's procedure will tell us that both are weighty. After all, being weighty is just a matter of not being undercut. But this isn't enough to settle how Ronnie ought to proceed.[177] To decide whether to go to the party, he needs to know something about how his reason to write compares with his reason to dance.

The source of the problem here is the recursive structure of Schroeder's definition. A reason is either undercut by some other set of reason or it isn't, so Schroeder's method can only produce binary judgments about the weight of reasons: weighty or not weighty. But this means that even if we answer the question about what the right kinds of reasons to undercut are, Schroeder still faces extensional worries in cases where an non-undercut agent-relative reason conflicts with a non-undercut agent-neutral reason. In at least some of these cases, the agent-neutral reason will seem weightier, and Schroeder will lack an explanation why.

One way for Schroeder to avoid this problem would be to conceive of reasons for one option as reasons against placing weight on reasons for another. For example, maybe Ronnie's reasons to write are ipso facto reasons not to place weight on his reasons to dance. If so, the structure of the case looks more amenable to his recursive strategy. But, I don't think it's plausible that all competition between reasons can be understood in terms of undercutting defeat. Suppose that a doctor promises to meet a friend for lunch, but on the way there encounters someone in dire need of medical help. Plausibly, the doctor's reasons to help compete with her

177 Thanks to Alessandra Yu for this example.

reasons to honor her promise. But is it also the case that her reasons to help undercut putting weight on her reasons to honor her promise? It seems not. Her reasons to help may be stronger, but they don't diminish the strength of her reasons to keep her promise. After all, if she does choose helping over promise-keeping, her reasons to keep her promise are still relevant: she'll owe her friend an apology or at least an explanation. So, this case seems to be one where reasons for an option cannot be redescribed as reasons for not placing weight on other reasons.

As a result, I think, Schroeder ends up in much the same position as the other internalists we've considered: making some progress on extensional worries, but ultimately not answering them. If so, it looks like purebred versions of both voluntarism and internalism have a hard time getting the extension of reasons right.

## 4. Taking Stock of the Prospects for Internalism

It seems, then, that even in sophisticated versions like the ones we've considered, internalism faces some extensional problems. Where does acknowledging this leave us? Given the general considerations about eliminative and non-eliminative reductions that started this chapter, internalists seem to run the risk of providing an analysis of reasons that if true shows there are no reasons at all. That is, they make talk about reasons too much like talk about the monster under the bed.

This result might seem unfair to internalists. One might think that, even if the best version of internalism doesn't allow us to accommodate *all* of our judgments about reasons, this shouldn't be grounds for rejecting the theory. Why not revise our judgments about reasons, instead of rejecting internalism? Part of the answer to this question is that the expectation that internalism

accommodate our intuitions about reasons wasn't just an abstract point about theory choice, where fit with intuition is one among several theoretical virtues. Instead, that demand was justified by reflection on what would make internalism count as a successful reduction, one that's not guilty of just changing the subject. Nonetheless, if internalists could identify grounds for putting less weight on some of the intuitions they have trouble accommodating, that would diminish the force of the arguments I've made so far.

I can think of two ways they might try to do so. One option would be to point out that some of the cases wielded against internalists are pretty weird, perhaps so weird that we should not be confident in our intuitions about them. After all, one might think, our intuitions are trained on ordinary life, and so might just not be well-equipped for evaluating characters like Perfectly Consistent Caligula.[178] I'm broadly sympathetic to this line of thinking, but I don't think it gets internalists off the hook here. For one thing, it's not the case that all the arguments I made above hinged on fantastical cases. The issue for Schroeder, for example, could come up any time there's a case of conflict between agent-neutral and agent-relative reasons. Similarly, my worry about Velleman and the lazy agent is – for this graduate student at least – all too recognizable.

To be fair, I have relied somewhat on extreme cases: the Mafioso problem for Korsgaard and Markovits, for instance. Still, if we imagine a real life version of someone like this – someone who deeply wants to be a Mafioso, but also has a normal range of prudential and familial concerns – I think the arguments still go through. While internalists might be able to appeal to this person's other concerns as a way of addressing the Too Few Reasons problem, the

178 See Street (2009) for some discussion of the problems we might have imagining cases like these. And see Railton (2014) for a very plausible account of intuition with the consequence that normative intuitions will be more reliable in more familiar situations.

Too Many Reasons problem looks just as daunting. We might think that this desire can be idealized away after Railton-esque idealization, but it's hard to see why. Maybe, after idealization, this person will shed their other concerns instead. If the internalist is right that we should be cautious about unfamiliar cases like this, presumably we should be cautious when speculating about what happens after idealization as well.

The other strategy is to try to restrict the range of cases that internalists need to accommodate. For example, Velleman is committed to prioritizing intuitions about action theory over intuitions about what we have reason to do.[179] If so, then it's not so bad if internalists have problems accounting for intuitions about ethics, so long as they can preserve intuitions about what's involved in acting for reasons in other domains.

Unfortunately, Velleman doesn't say much about *why* action theory should get priority here, but two rationales suggest themselves. One possibility is that the questions about action theory are somehow more fundamental; everything else we want to say about reasons is downstream from them. This may be true – I'm not sure – but I don't think it supports giving any extra weight to intuitions about action theory. We often find out how something works by looking at what happens downstream from it. For instance, weather patterns depend on atmospheric climate, but we can learn something about what's going on in the atmospheric climate by looking at what's going on with the weather patterns. So, concerns about fundamentality don't look like the sort of thing that could justify Velleman's approach. Another possibility is that questions about action theory are just the questions that happen to animate

179 See Velleman (1997) p41-42

Velleman.[180] If so, it'll be easier for him to revise his commitments in other areas of philosophy than in action theory. But it's pretty clear, I think, that this kind of rationale only explains why *Velleman* should prioritize intuitions about action theory, not why anyone working on the metaphysics of reasons should.

Another version of this strategy would be to provide some sort of debunking explanation for the intuitions that internalism has a hard time accommodating. Perhaps, for instance, our intuition that agents have reasons to be moral even when these reasons are not supported by their desires is just a function of our Puritan heritage or something along those lines.[181] If so, perhaps all internalists need to do is to accommodate the intuitions that would be left standing, once we get enough critical distance from our socio-cultural situation. In one way, I think this defense of internalism is on to something: I would sleep better at night if I had more to say about when and why philosophers should take their intuitions as starting points. Nonetheless, I don't think the mere possibility that a good story about that could vindicate internalism is enough to do the job here. After all, it might be that we should be discounting intuitions about the normative significance of desire because they're shaped by some self-interested or hedonistic aspects of our culture. Without hearing what the metaphilosophical story about when to trust intuition is, it's hard to know which intuitions it's going to leave standing. So, absent such a story, I think we should take the extensional problems with internalism at face value.

In short, I'm not sure that there's a good case for putting less weight on intuitions about

180 When I've chatted with him about metaphilosophical stuff, it seems like this is what he wants to say. But, these were informal conversations, so I'm not sure if this is his considered view and so don't want to pin it on him here.

181 Thanks to Dick Arneson for pressing me on this. I say "or something along those lines" because I take it this objection doesn't depend on any particular claims about sociology or religious history, just the general observation the intuitions have are susceptible to sociological explanation.

the extension of reasons. If that's right, then this way of objecting to the arguments I've developed here is at least unmotivated. Moreover, it's not just opponents of reductive views that want to take extensional intuitions seriously; tremendously clever defenses of internalism like Schroeder's and Markovits' proceed by trying to show how internalism can accommodate these intuitions. And this approach makes sense: part of what we want to do as philosophers is to systematize our thinking, resolve apparent inconsistencies, and see how things hang together. So, while I don't have a knockdown argument in support of proceeding this way, I think doing so should be relatively uncontroversial.

Moreover, even if we grant that one of the strategies discussed here can get the extension of reasons right, one might raise further worries about whether internalism gets the cases right *in the right way*. For example, one classic objection is that internalism makes our reasons out to be be too narcissistic.[182] When I make you a promise, it seems that my reason to keep it seems to be something primarily about you (e.g. your rights, legitimate expectations, or authority to call in this favor) rather than something primarily about me (e.g. my desire to keep my promises, or not to disappoint you). But if all of our reasons are a matter of our desires, internalism seem to have a hard time accommodating this observation.

Schroeder replies to this worry by noting that internalism doesn't have to say that the *content* of our reasons always has to do with our desires; our desires are just background conditions that explain why those considerations are reasons.[183] So, the content of my reason to drive you to the airport can just be "I promised to take you to the airport," even if the

182 See Brink (2008) for a particularly thorough development of this objection.
183 Schroeder (2007b), Chapter 2.

metaphysical explanation of why that consideration is a reason depends on my desires. Similarly, inauguration ceremonies might be part of the explanation of why someone is chair of the department, but it's not the case that the inauguration ceremony is part of the chair. Still, I think this response just moves the bump in the rug. Just as one might worry that internalism makes the content of our reasons objectionably narcissistic, one might also worry that it makes the metaphysical grounds of our reasons objectionably narcissistic. One might think, for instance, that it would be better if we sometimes acted for reasons that did not have a narcissistic ground. If so, other things equal, we should prefer a theory that does not force us to deny this. Viewed this way, Schroeder's move trades a problem with internalism's account of the content of our reasons for a problem with internalism's theoretical appeal.

To sum up: internalism's reductive ambitions put pressure on it to accommodate intuitions about the extension of reasons, and – even after canvassing various accommodation strategies – it's hard to see how it can do the trick. One might try to avoid this conclusion by taking on some metaphilosophical commitments about which kinds of intuitions theories about reasons ought to accommodate or what counts as accommodating them, but these moves look unmotivated. This might not rule out internalism – as with purebred voluntarism, there may be a way to get the cases right that we just haven't thought of yet – but it does put pressure on us to look for views that can do better.

# Chapter 4

## A Qualified Defense of Externalism

The question I'm trying to answer here is: in virtue of what does a given consideration count as a reason for action? So far, we have been looking at views that offer answers to this question in strictly non-normative terms. Facts about what agents will or desire are just descriptive facts about those agents, so if we can explain reasons in terms of these facts, we will have explained something normative in terms of something non-normative.

Externalism about reasons, by contrast, offers a normative answer to our question. For externalists, normative facts are not explicable in non-normative terms. If these normative facts are about reasons, then the answer to our question will be something like "that's what the normative facts are." If these facts are about some other normative property (say, goodness or virtue), then the answer to our question will be in terms of that other normative property (say, that acting on this kind of consideration would be good, or that virtuous agents would act on it). Either way, the thing in virtue of which a given consideration is a reason will be a normative thing. In this way, externalism is a "realist" view about the normative; it holds that normative claims are truth-apt, some of them are true, and that when they are true they are true in a mind-independent way.[184] On these views, normative facts are a fundamental commitment of our

184 This characterization is a bit rough. To make it more precise, one thing we'd need to know is what kind of mind-independence a view needs to recognize in order to count as realist. Consider, for instance, Korsgaard's voluntarism. According to Korsgaard, there are reasons we all have just in virtue of being agents, because our reasons depend on our wills and there are some things all agents must will in order to count as agents in the first place. Are these reasons mind-dependent? In one sense, yes, because they depend on mental facts about the agents who have them. In another sense, no because they don't depend on facts about any *particular* mind, just facts about agency considered abstractly. So, for instance, Korsgaard can say that there are facts about my reasons that are independent of what I think about them. Does this make Korsgaard's view a realist one? In Korsgaard (2009), she is ambivalent. It seems to me that whether the label applies depends on how robust a notion of mind-independence you have in mind. To avoid getting into the fray about who counts as a realist, I'll stick to the "externalist" label here.

ontology.[185]

In characterizing externalism this way, I hope to remain neutral on what externalists say about the *content* of the normative facts (though I'll have a more to say about the options later in this chapter). Instead, I'm making a claim about what externalists say about the question: in virtue of what are those contents normative? In particular, I don't mean to rule out that some or all of the normative facts might in some way make reference to agents' mental states. For instance, it might be that one of the normative facts is that it's good to experience pleasure, that it's good to have your desires satisfied, or that it's wrong to hurt others' feelings. Externalists can recognize such facts – at the extreme, they could hold that all normative facts are like this – without also saying that our pleasures, desires, or feelings are what metaphysically explain why some considerations are reasons.[186]

In any case, this non-reductive feature of externalist views may seem to give purebred versions of voluntarism and internalism a structural advantage in the debate about how to answer

185 Though, as I'll explain below, exactly what kind of commitment ontological commitment externalists are making is a matter of intra-externalist dispute.

186 A comparison might be useful here. Suppose that I want some coffee. Internalists and externalists can agree that the consideration "there is coffee in the lounge" is a reason for me to go to the lounge. They can't agree, though, that my desire is the thing in virtue of which that's a reason. Internalists will say that, while externalists will say it's that plus the mind-independent facts about reasons, goodness, or some other normative concept. The same point applies to pleasure. Suppose I adopt a first-order theory according to which the only reasons I have are reasons to pursue pleasure. In adopting this theory, I haven't yet chosen sides between internalism and externalism. To do that, I'd have to say in virtue of what I have those reasons. Is it my desire for pleasure, the goodness of experiencing that mental state, or what?

This is not to say that all combinations of views are equally plausible. If one thinks, as a matter of first-order theorizing about what reasons we have, that all reasons are a matter of desires or pleasures, then it might seem that one might as well go in for metanormative internalism. After all, as we saw in Chapter 3, the main objections to internalism hinge on denying these judgments about what reasons we have. If you're already biting those bullets, you might as well get the payoffs internalists advertise (e.g. accommodating a tight connection between motivating and justifying reasons, or doing without externalists' ontological commitments). But this is a point about which combinations of views are most plausible, not about which combinations of views are logically possible.

our question about the metaphysics of reasons. *If* they can get the extension of reasons right, then their theories get by with fewer fundamental ontological commitments. But, as I've argued over the course of the previous two chapters, they do not get the extension of reasons right. These extensional difficulties for purebred voluntarists and internalists constitute a *pro tanto* argument for externalism. There are some facts about reasons – say, that a committed Mafioso does not have decisive reason to kneecap her rivals – that externalism gets right while these other views get wrong.

Still, this is just a *pro tanto* case. It might be that there are other costs to recognizing externalist facts. If these costs are high enough, then perhaps we ought to revise our views about the extension of reasons rather than accepting externalist facts. In this chapter I'll argue that this is not the case, and so the *pro tanto* argument for externalism holds up.

Nonetheless, I mean this only as a qualified defense of externalism. As I see it, the conclusion of the *pro tanto* argument is only that in order to avoid extensional trouble, a view about the metaphysics of reasons ought to recognize some normative facts that are not explicable in non-normative terms. It does not follow from this that the best view about the metaphysics of reasons is purebred externalism; hybrid voluntarists recognize some normative facts that are not explicable in non-normative terms too. If, as I'll argue in Chapters 5 and 6, hybrid voluntarists have other, non-extensional advantages over purebred externalists, then we should opt for a hybrid voluntarist view. So, my goals in this chapter are somewhat modest. I'll be arguing only that the standard objections to externalism do not rule out views that recognize externalist facts; I won't be arguing that, all things considered, purebred externalism is the way to go, or even that

recognizing such facts doesn't come with any theoretical cost – just that if there are costs, they aren't the kind that rule out externalism the way extensional difficulties rule out purebred versions of internalism and voluntarism.

## 1. Varieties of Externalism

Before we can tally up theoretical costs and benefits, though, we need a more thorough characterization of what externalism actually says. So far, I've characterized it in purely metanormative terms: it's the view that when we answer the metaphysical reading of the question "in virtue of what is this consideration a reason?" we appeal to some facts which themselves have normative content. But a view that says that there are normative facts isn't complete without saying something about the content of those facts. On this score, I think, externalists have many plausible options, so we should turn our attention to those.

As I noted above, one choice for externalists is which normative concept they think we should appeal to when we answer questions about in virtue of what some consideration is a reason. Perhaps considerations are reasons in virtue of what's objectively good for agents, or objectively valuable, or what it'd be objectively virtuous to do. For the sake of simplicity, though, I'm going to set this question aside and focus on externalist views where the fundamental normative facts are about reasons. This is just for ease of expression; I think what I say could be translated if need be.[187] So, we can now recast the question about the content of the normative facts as the question: what do agents have objective reason to do?

One way to answer this question would be to provide a list. Such *objective list* views

187 See the Postscript to Chapter 1 for more on this choice.

offer a tidy answer to the extensional problems I pressed for internalists in Chapter 3. Why doesn't Perfectly Consistent Caligula have a reason to cause harm? Causing harm isn't on the list. Why does Scrooge have a reason to give to charity? Helping those in need is on the list.

Of course, there are different views one could take about what exactly goes on the list, and so there are many more specific views in the objective list family. Common choices include things like experiencing pleasure, appreciating beauty, developing friendships, caring for others, and achievements in intellectual, aesthetic, or athletic pursuits. The point for our purposes here, though, is that whatever the precise content of the list, this is one general form that externalist views can take.[188]

Objective list views offer straightforward answers to questions about the extension of reasons, but this straightforwardness might seem to come at the cost of making them less theoretically satisfying. In particular, they seem to preclude our having answers to questions about *why* some item is or isn't on the list. While externalists' answer to a metanormative reading of this question (i.e. what metaphysically explains why this item is on the list) has to be something like "that's what the normative facts are," this needn't preclude them from offering an informative answer to a normative reading of the question (i.e. what justifies this item being on the list). In other words, they can still have something to say about what reasons we have for caring about or pursuing the items on the list.

This impulse to provide some further theoretical structure lends support to *perfectionist* versions of externalism. Perfectionist views pick out a salient kind of thing we are, like that we are humans or agents, and then go on to explain what we have reason to do in terms of

188 See Arneson (1999) for an exploration of the appeal of these views.

developing or achieving the excellences for that kind of thing. One way to think about perfectionist views, then, is that they offer an *organized list* answer to what we have reason to do, where the organization comes from each item's contribution to perfection for members of a salient kind. This organization gives perfectionist views the resources to say more than bare objective list views can about why any given item makes it on or off the list.

This is an abstract characterization, so it might help to have an example. One important contribution to the perfectionist tradition comes from T.H. Green. On Green's view, what's importantly distinctive about humanity is our ability to deliberate. We can get critical distance from our desires in order to evaluate them, and we can form new desires on the basis of our reflection about the good. For, Green, then, what humans ultimately have reason to pursue is the development and use of these deliberative capacities. He goes on to argue that this kind of self-realization requires harmonizing one's ends with others, such that we all realize our nature as agents together.[189]

For the purposes of this chapter, I think we can stay neutral on which of these ways of developing externalism is more plausible. On the one hand, perfectionist views seem to offer more explanatory power, at least at the normative if not metanormative level. On the other hand, they run the risk of ending up with less tidy answers to questions about the extension of reasons. For perfectionist views to get the extensional questions right, they'll need a story about why each of the things we have reason to do is relevantly connected to membership in the relevant kind. As we saw with Korsgaard in Chapter 2, spelling out these kinds of explanations is no easy task. For

189 See Green (1883) for the full story and Brink (2003) for a sympathetic reconstruction. See Hurka (1993) for a different way of developing the perfectionist strategy.

instance, on perfectionist views, it might be hard to see how to explain the reasons agents have related to what Arneson calls “cheap thrills,” i.e. low-cost pleasures that don't contribute in a meaningful way to our development. That something would be a cheap thrill does, at least often, seem like a reason to do it, even if doing it doesn't express my deepest nature.[190]

Nonetheless, either way of developing externalism will have answers to questions about the extension of reasons that are not available to reductive views like purebred internalism and purebred voluntarism. Both kinds of externalists can point to something mind-independent (either the list, or the perfections for members of the relevant kind) to explain the presence or absence of reasons in the cases that raise trouble for purebred internalism and purebred voluntarism.

So, recognizing externalist facts can help us make progress in vindicating our judgments about the extension of reasons. We now need to ask, what might be objectionable about recognizing these kinds of facts? Broadly speaking, I see four candidate objections. It might be that externalism commits us to a spooky or otherwise non-parsimonious metaphysics, it might be that externalism fails to capture the broadly affective features of reasons like their role in motivating agents or what it feels like for agents to deliberate, or it might be that it introduces epistemological puzzles about how agents can be justified in their judgments about what reasons they have. In the remainder of this chapter, I'll take these objections in turn.

190 I'm not sure what to make of this objection. How plausible it is depends, I think, on what the perfectionist says the excellences for creatures like us are. Maybe what's distinctively human isn't just rational agency, but also things like experiencing thrills, cultivating friendships, etc. Or maybe there's a conception of rational agency that packs those kinds of things in.

## 2. Objections about Metaphysics

Let's start with metaphysics. I observed above that externalists take on board an ontological commitment that internalists and purebred voluntarists do not. There is a rich tradition of objecting to externalism on the grounds that there is something spooky or otherwise objectionable about this extra ontological commitment.[191] So, to start with, we should see if we can hone in on what kind of commitment is being made and what exactly is supposed to be objectionable about it.

The objection cannot be that there are some facts that externalists leave unexplained. After all, every theory turns its spade somewhere. For example, suppose than an internalist holds that desires are a fundamental psychological kind, not reducible to others. This might or might not be the right account of the psychological facts, but it wouldn't give rise to the metaphysical worries philosophers have about externalism. So, if there's something objectionable about irreducibly normative facts, it's not just that they're irreducible.

A more promising objection relies on a comparative judgment: externalists turn their spade *earlier* than internalists or voluntarists do. Given some normative fact that externalists take to be explanatory bedrock (say, that some consideration is a reason), internalists and voluntarists will have something more to say about in virtue of what it's true (say, that an agent has willed in such-and-such a way). But this can't be the whole of the objection either, because ontology isn't the only dimension along which we assess theories. Recognizing more fundamental ontological kinds is only objectionable if you don't need those kinds to explain the phenomena, and the extensional problems for internalism and voluntarism strongly suggest that we do.

191 Mackie (1977) got this ball rolling.

To make sense of the ontological objection to externalism, then, we need to identify something distinctive about *normative* facts that would make recognizing them as a fundamental ontological kind objectionable, rather than just any new ontological commitment. I think there are two possibilities.

One worry is that recognizing irreducibly normative facts doesn't just commit us to a new ontological kind, but a different sort of ontological kind. In particular, one might worry that normative facts aren't *natural* or compatible with a broadly-speaking modern, scientific worldview.[192] Consider a physicist who discovers a new kind of particle. We might think "Sure, we're adding to our list of fundamental commitments, but we already have *physical stuff* on the list broadly, so adding one more kind of physical stuff isn't that big a deal." Contrast this with the suggestion that we add normative facts to our ontology; in that case, it's harder to see the addition as continuous with previous entries on the list.

Responding to this objection is a complicated task, in part because defenders of externalism differ about exactly how to understand their metaphysical commitments. Some externalists take the non-naturalness intuition on board, and argue that the bedrock normative facts really are discontinuous with ordinary empirical ones. On these views, we're supposed to think of normative facts as something like Platonic forms.[193] Nonetheless, defenders of these views argue, we should not think that an ontological commitment has to be a natural one in order for us to be justified in accepting it. Other externalists offer ways of thinking about the normative that make it less discontinuous with the non-normative. For instance, it might be that as a matter

192 Admittedly "modern, scientific worldview" isn't a particularly illuminating description. See Sturgeon (2007) p64 for a quick survey of what this idea has been used to mean in this context.

193 Enoch (2011a) defends this kind of view. In fact, he leverages the thought that normative facts seem different in this way into an argument against other forms of realism.

of analytic fact, particular bits of normative language (like "reason") just mean the same thing as particular bits of descriptive language. If so, using those bits of language wouldn't amount to making any new ontological commitments. Another possibility is that there are synthetic facts about the relations between normative and non-normative kinds. Perhaps particular normative kinds stand to particular descriptive kinds the way water stands to H2O.[194] Alternatively, externalists might deny that any identity relation holds between the normative and the descriptive, but maintain that there is some other relation between them which makes normative facts respectable on a naturalistic worldview. Perhaps, for example, the normative is *grounded in* or *constituted by* the natural, so that normative facts stand to descriptive facts the way statues stand to the clay they're made of.[195] Statues aren't identical to clay (you could, say, destroy the statue without destroying the clay it's made of) but there's nothing metaphysically spooky about the existence of statues given the existence of clay.

Things are further complicated by the fact that what it takes to be a "natural" fact in the first place is also a matter of philosophical dispute.[196] So, rather than make an argument that depends on the particular details of these various proposals, let me make a broader dialectical point about what's at issue between externalists of different metaphysical stripes and their critics who worry about naturalness.

These different versions of externalism have their costs and benefits for us to tally up. Once we do, maybe one according to which normative facts are natural will be clearly best, or

194 See Shafer-Landau (2003) for discussion of these possibilities.
195 For an example of this strategy, see Brink (1989). Another possibility, in a similar spirit, comes from Stringer (2018). On Stringer's view, normative facts are emergent properties.
196 See Sturgeon (2007), Shafer-Landau (2003), Copp (2003), Ridge (2007), Enoch (2011a), and Cueno (2007) for some discussion.

maybe accepting one like that would require biting some metaphilosophical bullets about what makes a commitment naturalistically respectable. For example, we might revise our conception of naturalness, or our methodological assumptions about whether naturalness is necessary for metaphysical respectability. Nonetheless, if – as I've argued over the last several chapters – we need externalist facts to get the extension of reasons right, then worries about the naturalistic credentials of normative facts can't on their own be sufficient for rejecting externalism. Instead, at best they are sufficient for *either* rejecting externalism *or* biting whatever bullets one needs to bite in order to make normative facts come out non-spooky. So, if this naturalism objection is going to convince us not to recognize externalist facts, it needs to come supplemented with an argument for going that way rather than opting for a version of externalism according to which normative facts are natural ones or where it's okay to have some non-natural facts in our ontology. But, given the extensional problems for views that don't recognize externalist facts, there is strong pressure for going the latter route. There are some extensional claims about reasons – say, that I have a reason to avoid pain – about which I am more confident than I am in any metaphilosophical claim about what naturalness is or its relation to ontology.

This puts me and an objector who worries about externalist facts' naturalist credentials in something of a "one philosopher's *modus ponens* is another philosopher's *modus tollens*" situation. If offered an argument from a conception of naturalness to the claim that externalist facts would commit us to an objectionable metaphysics, I will be inclined to interpret it as an argument against that conception of naturalness, or against using that conception of naturalness as a guide to metaphysics, rather than against recognizing externalist facts.

This way of framing things might sound objectionably dogmatic. In some sense, I'm just insisting that *something* must ground facts about reasons, and that might be unsatisfying compared to offering an explanation of those facts in non-normative terms. There is something to this worry, I think; I really would prefer a theory that could pull off that trick. But, I've been arguing over the last several chapters, those theories don't deliver the goods. So, while it would be dogmatic to start from the assumption that recognizing externalist reasons is the only way to vindicate our intuitions about reasons, it's not dogmatic to accept this on the basis of arguments against competing views.

This all suggests that for our purposes here we can safely set aside questions about exactly how to figure out when an ontological commitment is a naturalistically respectable one, how best for externalists to meet that standard, or whether they even ought to. Nonetheless, distinguishing the natural and the non-natural isn't the only way of developing the thought that there is something distinctively odd about an ontological commitment to normative facts. One might instead worry about those facts' *explanatory role*.[197] Suppose, for instance, that one of the bedrock normative facts is that everyone has a reason to take the means to their ends. Suppose further that this fact is a natural one, whatever that turns out to mean. Still one might wonder whether we need to appeal to this fact in order to explain anything that happens in the world. For instance, one of my ends is completing a dissertation, and a necessary means for this is setting aside time to write. But when I do so, it's not clear that one needs to appeal to facts about reasons at all in order to explain what I'm doing. Instead, we might get by with purely psychological descriptions, like about what I want to do and what I believe about how to do it. Similarly, given

197 Harman (1977) raises this issue. See Chapter 1, "Ethics and Observation."

that people can fail to do what they have reason to do, we might wonder what possible behavior of mine could cast doubt on our confidence in our beliefs about what reasons I have. If such facts are immune from empirical challenge, and aren't needed to explain what happens, why believe in them?

One might resist this objection in several ways. First, it's not so clear that normative facts don't have a role to play in explaining non-normative phenomena. After all, even when we explain non-normative phenomena in terms of non-normative facts, we do so using bridge principles about the relations between various facts. For instance, if we explain my setting aside time to write in terms of my desire to complete my dissertation and a belief about how to do so, we can do so only in conjunction with some background assumptions about how beliefs and desires move agents to action. So, it's not clear why normative facts couldn't avail themselves of comparable background assumptions.[198] For example, if your roommate promises to do the dishes and then does them, we might explain her doing them by appealing to the normative fact that she has a reason to keep her promises plus the background assumption that she's a considerate person who reliably responds to promise-keeping reasons.

Second, it's not clear that normative facts need to earn their keep by explaining non-normative phenomena. After all, there are normative phenomena to explain.[199] Take the fact that a committed Mafioso does not have decisive reason to shoot her rivals' kneecaps. This is a fact that calls out for explanation, and one that purebred voluntarism does not seem to be able to explain. By contrast, externalism does have an explanation for it, one in terms of the normative

198 See Brink (1989) p182-197 for discussion of this strategy.
199 Nagel (1970) discusses this possibility.

facts. So, if ontological commitments earn their keep by explaining things that we otherwise wouldn't be able to explain, externalist facts seem to pass this test even if the things they explain are not empirical.

This situation we're left with is the following. Many philosophers find it hard to believe that there could be irreducibly normative facts. I myself only came to believe in them after not being able to come up with an extensionally adequate version of purebred voluntarism. Nonetheless, it's hard to articulate what's so implausible about recognizing these kinds of facts in a way that starts with methodological assumptions that aren't themselves open to question. So, I suspect, what's going on here is that worries about externalism committing us to a spooky metaphysics are really stand-ins for other kinds of objections. They're either motivated by the conviction that purebred versions of reductive theories really can be made to work, or the thought that there is something important about reasons that externalists have a hard time explaining. If the former, a commitment to externalism would be unmotivated. If the latter, we might think that externalists' metaphysical commitments don't really earn their keep. Since I've argued against the former possibility in the previous two chapters, we should now turn our attention to the latter.

## 3. Objections about Epistemology

The same kinds of considerations that might make irreducibly normative facts seem metaphysically objectionable can also give rise to epistemological concerns. If normative facts are in some way discontinuous with the descriptive ones, one might wonder how we come to be justified in our beliefs about them. After all, our ordinary senses and reasoning practices seem

geared to tell us how the world is, rather than how it ought to be. One can, for example, use sight to ascertain that your brother cut himself a bigger slice of pie, but sight on its own won't tell you that the pie distribution was unfair, or that your brother had a reason to leave you more than he did.

A natural first pass at allaying these concerns points out that much of our reasoning about how the world is doesn't rely on direct observation in the way that examples like sight suggest. Much of our reasoning proceeds by trying to systematize and extrapolate from a set of beliefs already in place. Even cases of more-or-less direct observation depend on considerations of coherence with what we already think: if my visual experience is sufficiently weird, I'll take myself to be hallucinating rather than seeing. So, one way for our normative beliefs to be justified would be for them to cohere well, and one way to acquire this justification is to try to bring them into coherence, e.g. by adjusting big picture principles in light of judgments about particular cases, and vice versa.[200]

Of course, coherence isn't everything. One might have a coherent set of beliefs about the Tooth Fairy – when she visits, the going rate for teeth of various kinds, what motivates her to collect teeth in the first place – but this is no reason for thinking that the Tooth Fairy actually exists. So, we can't answer worries about how we acquire justification for our normative beliefs merely by appeal to having brought those beliefs into reflective equilibrium.

This point helps sharpen the possible epistemic worry about externalism. For our empirical beliefs, coherence seems justifying in part because we are confident that the starting points in our process for making our beliefs coherent are at least on the right track. I trust the

200 Rawls (1971) calls this strategy "reflective equilibrium."

output of inductive reasoning in part because I trust the particular judgments that form the inductive base. And, though there is philosophical controversy about how best to understand them, our starting points for empirical beliefs – things like our senses and memories, and the testimony of others – all plausibly seem like ways of getting a grip on what the world is like.[201] But, if normative facts are not the kind of thing we can find out about this way, then it's harder to see why we should be confident in them as inputs into reflective equilibrium. Or at least, it's harder until we find some reason for trusting the starting points for reasoning about the normative.[202]

Many philosophers have proposed answers to this question, and for my purposes here I don't want to take sides about exactly what the best view is (like the question about the best version of externalism's metaphysical commitments, this could be its own dissertation). Nonetheless, I'll sketch an approach that I find plausible. This should at least take some of the sting out of epistemic worries about externalism; it should move an objector from thinking externalism is a non-starter to thinking that it might require biting some bullets, but none so hard to bite that we shouldn't accept it given the extensional problems with other views.

The question the externalist faces here is: why treat our intuitions about the normative as appropriate starting places for inquiry? Enoch points out that this question looks like a special case of a more general kind of question: why treat any belief-forming mechanism as an

201 Kelly & McGrath (2010) press this objection in a particularly clear and forceful way.

202 One way to reply to this challenge might be to shift the burden back on to the objector. One might say, "look, give me an argument for not trusting my normative beliefs." Ultimately, I think there's something right about the idea that we should put some stock in our normative beliefs unless we're given a reason not to, and I'll develop that strategy below. But, to be fair to this kind of objector, I don't think this epistemological challenge to externalism needs to depend on any further argument meant to cast doubt on the starting points for normative inquiry. Beliefs call out for justification in a way that the suspension of belief does not, so it doesn't seem like a stretch to think that the burden of proof is on those of us who think we're justified in our normative beliefs.

appropriate starting place?[203] Once the question is posed in this general way, it's clear that answering it can't always be a matter of identifying a further source of justification. Some belief-forming mechanisms, like trusting our perceptual seemings or reasoning by *modus ponens*, are *basic*, in that we are entitled to rely on them without being able to explain their justifying power in terms of some further mechanism.[204] So, rather than looking for some other source of justification that might vindicate our normative intuitions, Enoch thinks, we ought to be asking whether those intuitions themselves are plausibly one of the basic belief-forming mechanisms.[205]

Of course, it would be no good just to assert that our normative intuitions are a basic belief-forming mechanism. Reasoning by *modus ponens* is one, but reasoning by affirming the consequent isn't. Similarly, reasoning by induction is, but reasoning by counter-induction isn't. Trusting our perceptual seemings is, but there's room to wonder whether trusting our normative intuitions is too. So, to make good on the idea that our normative intuitions might be basic in this way, we need to say something about what makes a belief-forming mechanism a good candidate for being a basic one. Since these are supposed to be basic methods of justification, we can't do this by finding some further thing that justifies some of these methods but not the others; such a justification would make them non-basic. Nonetheless, we can at least try to find out what the seemingly good methods (*modus ponens*, induction, perception) have in common that's not

203 One might worry that this strategy doesn't address the full generality of possible epistemological problems for externalism. See Street (2006) for an argument that, even if we have a justification for in treating our normative intuitions as an appropriate starting place, there are always defeaters for that justification. In future work I hope to have more to say about this kind of challenge, but for my purposes here I'll focus on the more basic worry about whether we can get an epistemic grip on the normative in the first place. At any rate, many of the responses to Street in the literature draw on the idea that we have at least *pro tanto* justification for normative beliefs, so establishing that at least a start toward a reply to Street. See Locke (2014) for discussion.

204 This is not to say that these mechanisms are infallible, just that we can put some *pro tanto* stock in what they tell us without having some further justification for doing so.

205 See Enoch (2011a). Chapter 3 develops this strategy.

shared by the bad ones (counter-induction, affirming the consequent) so that we can use this description to test controversial cases like normative intuition.

A natural starting place here is the idea that some belief-forming mechanisms are *necessary* for us. Without trusting our perceptual seemings, for example, we'd have no hope of having justified beliefs about the world around us. Similar suggestions are sometimes made in defense of our intuitions about mathematical facts.[206] But, Enoch points out, necessity for some kind of inquiry can't be sufficient for our putting default stock in a belief-forming mechanism. Reasoning by counter-induction might be necessary for something (say, winning the bad inference championships), but this doesn't show that we ought to treat counter-induction as a basic belief-forming mechanism. Enoch proposes to refine this idea by adding the condition that it's *necessity for a valuable project* that's distinctive of basic belief-forming mechanisms. We could give up trying to learn about the external world, but that would come at a huge cost.

It also matters, Enoch thinks, how the mechanism does in practice. Imagine, for example, that we lived in a lawless world where induction didn't help us. In this world, learning about the external world would still be a valuable project, but induction wouldn't give us hope of success at it. In these circumstances, we wouldn't be justified in treating induction as a basic belief-forming mechanism. The same applies to a world where our perceptual seemings don't help (e.g. one where we hallucinate regularly and convincingly).

While considerations about necessity for inquiry are often wheeled in in support of belief-forming mechanisms that are necessary for scientific inquiry, in principle the same kinds of things might be said about normative inquiry. Just as we can't inquire about what is without some

206 Enoch (2011a) p54-57

way of getting a grip on the empirical facts, we can't inquire about what to do without some way of getting a grip on the normative facts. So, the epistemic question facing the externalist is whether trusting our normative intuitions meets the same conditions our basic mechanisms for forming descriptive beliefs do.

To get at this question, we can apply Enoch's proposal: a belief-forming mechanism is appropriately treated as basic if and only if (a) it's necessary for some project, (b) the project is valuable, and (c) there is at least some hope of succeeding at the project using the method.[207] If this is on the right track, it looks like our normative intuitions are on the right side of the ledger.[208] With respect to (a), trusting our normative intuitions seems to be necessary for any forward-looking thinking about what to do. With respect to (b), this kind of deliberative agency is central to the kinds of creatures we are, and a prerequisite for anything else we might value doing.

One might worry, though, about how well normative intuition satisfies condition (c). After all, there is pervasive disagreement about the normative, both with respect to particular cases (e.g. about our duties to distant needy persons) and theoretical questions (e.g. about whether the demandingness objection is a serious issue for consequentialists). And unfortunately, Enoch doesn't say much about what successful inquiry in some domain means, and *a fortiori* what it would be to have a reasonable hope of success at it.

Nonetheless, I think some parallels with our paradigm basic belief-forming mechanisms can offer guidance here. Consider sight. Placing default trust in what one sees doesn't resolve all

207 Enoch (2011a) p83
208 Enoch (2011a) p83

disagreement about the observable world; you and I might make different judgments about the result of a close horse race. Nor is sight perfectly reliable; we might agree about the horse race and just turn out to be wrong. So, hope of success here can't mean perfect reliability or the elimination of disagreement. Now consider inductive inferences. The valuable epistemic project that such inferences help us with – figuring out what's going on in the world around us – isn't one that we can *only* make progress on by using induction. We can make some progress on it by way of direct observation. So, the kind of hope of success required for being a basic belief-forming mechanism can't be being our only hope.

With these comparisons in mind, it looks much more plausible that normative intuition could satisfy condition (c). Thinking about reasons sometimes provides clear verdicts that can help to guide our actions at least some of the time, and that seems to be all that's needed to earn our default trust. So, if meeting these conditions is good enough for trusting our perceptual seemings, the same looks like it goes for the normative.

## 4. Objections about the Affective Features of Reasons

Our last category of possible objections to externalism focuses on the broadly affective features of reasons. I think there are two issues here. We can address the first fairly quickly. One might worry, like Korsgaard does, that externalism doesn't explain the distinctive authority of reasons, or how they get a grip on us first-personally. Appealing to something like an agent's practical identity might seem to do better at explaining what makes reasons feel authoritative for us when we're deliberating about them. As I argued in Chapter 2, though, voluntarism does not really offer an advantage over externalism on this score. The upshot of my discussion of Enoch's

shmagency objection was that it's hard to spell out this issue in any non-question-begging way, so externalists like Enoch can't use it as an objection to voluntarists like Korsgaard, nor can voluntarists like Korsgaard use it as an objection to externalists like Enoch.

The other thing an externalist account of reasons might seem to struggle to explain is normative motivation. Making judgments about what we have reason to do can sometimes motivate us to do it. Sometimes, I can get myself to go to the gym by thinking about the reasons I have to do so. But since externalism doesn't link the truth of our judgments about reasons to something in our motivational system, it might seem that externalists lack an explanation for this phenomenon.

Following Brink, we can represent this idea more formally by looking at the following claims, each plausible but mutually incompatible: (1) normative judgments express beliefs, (2) normative judgments entail motivation, (3) motivation involves a desire or other pro-attitude, and (4) there is no necessary connection between any belief and any desire or pro-attitude.[209]

Strictly speaking, this puzzle arises for all of the views we've considered so far. Externalism, internalism, voluntarism, and hybrids between them all offer competing explanations for what makes facts about reasons true, so, on all of these views there are facts about reasons and our judgments about them express beliefs about those facts.[210] But things may seem to be particularly pressing for externalists. To see why, contrast the situation for internalists. On their view, whenever an agent has a reason, this is made true by the agent's

209 Brink (1997a) frames the puzzle this way. He is specifically focused on moral judgments, but structurally similar issues arise for normative judgments more generally.

210 This is just to say that the views I've been considering here all accept this. Not all philosophers do. One might, for instance, think that normative judgments aren't a matter of belief but of some non-cognitive attitude. See, for example, Gibbard (2003) and Blackburn (1993).

having some desire. So, agents will have some motivational push in the direction of acting on their reasons, whatever they believe about them. In contrast, since externalists don't explain reasons in terms of something in our motivational system, if they are going to explain how reasons motivate us, they're going to have to do it through our beliefs about them.

To make that explanation work, it looks like externalists have to give up on at least one of the above claims. Brink's preferred way out of the puzzle is to deny (2); on his view, while there's no necessary connection between normative judgment and motivation, there is a reliable, non-accidental connection between on them. More specifically, he thinks that what explains why a given agent will be motivated by her sincere normative judgments is that she also accepts the authority of those judgments.[211] For instance, someone who is motivated by her judgment that morality requires donating to charity is so motivated because she accepts the normative authority of the standards of morality. Similarly, when I am motivated by my judgment that prudence requires exercise, I am so motivated because I accept the normative authority of the standards of prudence.

Other philosophers who want to hang on to (2) go other ways. For example, one might deny (3) by holding that some beliefs can motivate us without the help of the other attitudes.[212] Or one might deny (4) by holding that some beliefs necessarily produce certain desires in the agents who have them.[213] What each of these solutions to the puzzle have in common is a commitment to the idea that there is a reliable connection between judgments about what reasons we have and what we'll be motivated to do. These views differ on how tight the connection is (it

211 Brink (1997a) p30-32
212 For example, see McDowell (1978)
213 For example, see Nagel (1970)

might be merely reliable or it might be constitutive) and the mechanism that produces it (beliefs might motivate on their own or through connection with some other attitude), but each embraces the connection in some form or other. The upshot here is that although externalists do need a way out of this inconsistent quartet, the claims that generate it are stronger than they first appear. One can accept something close to them while still recognizing a relatively tight connection between judgments about reasons and motivation to act.

Still, one might think that there is more explaining for externalists to do. The connection that each of these strategies postulates – one between judging that you have a reason and being motivated to act on it – might itself seem like something that needs explaining. Why accept, for example, that some beliefs can motivate, or that some beliefs tend to produce certain desires in the agents who have them? Sure, one might think, if something like that were true it would show how normative motivation is possible, but why think that something like that is true?

A tempting way for externalists to answer this question is to make a constitutive claim about agency. Being responsive to your reasons is a prerequisite for being an agent at all, so it's not something that needs further explanation, an externalist might say. And as far as constitutive claims about agency go, this one is very plausible. Someone who isn't motivated by some of her judgments about reasons seems to that extent irrational, and someone who was never so motivated is hard to recognize as an agent.

I am of two minds about this reply. On the one hand, I think it's very plausible that being motivated by your judgments about reasons is part of rational agency. And in light of that connection, the phenomenon of normative motivation needn't be a puzzling one for externalists.

On the other hand, I share the objector's feeling that it would be nice if we could say more about *why* there is this connection. Pointing out that something is required by rationality is not always the same as identifying something that will motivate an agent to do it. We might be alienated participants in agency, or parts of agency. For instance, we might imagine someone who in general is motivated to act on her reasons but doesn't feel the tug of moral ones, or someone is motivated by reasons having to do with others' interests, but systematically neglects her own. It would be helpful to be able to say something about why such agents are the defective case rather than the default.

Jamie Dreier diagnoses this situation in the following way. While both externalists and their critics can agree that there *is* a tight connecting between being a rational agent and having some link between your cognitive system and your motivational system, one might still worry that externalists leave it mysterious what explains this connection.[214] And, if one is worried about explaining this, one won't be satisfied if one's interlocutor just points out that the claim one wants an explanation for is very plausible (compare: it's very plausible that we have reason to be moral, but this doesn't make arguing for moral rationalism a less interesting philosophical project). So, while existing solutions to the puzzle of normative motivation can help shed light on the *mechanism* that produces that motivation, there is room to wonder why being rational is (in part) a matter of having that mechanism, whatever it is.

We're left with this. I don't think externalism loses much plausibility for facing the puzzle of normative motivation; that there is some reliable connection between agents cognitive and

214 See Dreier (2015). Dreier offers this as a way of understanding Korsgaard's complaint about externalism in Korsgaard (1996). Thanks also to Dustin Locke for discussion on this point.

affective states needn't be mysterious. Still, I think there is room for hybrid views to gain plausibility by offering a deeper explanation of this phenomenon. So, I'll return to this idea in Chapter 6.

Let's step back a bit. This chapter has avoided a discussion of what the best version of externalism is; different strategies for addressing the objections we've considered come with different costs and benefits. But I do take the foregoing discussion to establish at least the following: the classic objections to externalism are best seen as offering ammunition for arguments about *which* version of externalism to accept, rather than about whether to accept it at all. For instance, concerns about metaphysics look like arguments for a naturalist version of externalism, while concerns about epistemic access to the externalist facts look like arguments for treating normative intuition as a basic belief-forming mechanism. So, however those arguments ultimately shake out, we can be confident that a theory of the metaphysics of reasons shouldn't be ruled out for accepting externalist facts.

That point, plus the extensional difficulties with purebred versions of internalism and voluntarism, suggests that the best view is either going to be a purebred externalist one or a hybrid with an externalist component. In the chapters that follow, I'll argue for the second possibility. In particular, I'll try to show that accepting a voluntarist component offers some explanatory advantages over a purebred externalist view.

# Chapter 5

## Hybrid Voluntarism, The Story So Far

The arguments I've made so far have been primarily negative.[215] Specifically, I've argued that purebred versions of internalism and voluntarism face extensional difficulties. If so, we ought to accept a view with an externalist component. The plan for the rest of the way is more positive: I'll argue that we should prefer a view that is externalist about some reasons and voluntarist about others over purebred externalism. In this chapter, I'll canvass some arguments from Chang and from Andrea Westlund, offering some friendly amendments and novel responses to objections. I'll also raise some questions about whether these arguments support Chang's particular version of hybrid voluntarism uniquely well, or whether the space of possible views is wider. In the next chapter, I'll offer my own independent arguments for hybrid voluntarism, and revisit the question about the most plausible version of hybrid voluntarism in light of them.

### 1. Classic Arguments for Hybrid Voluntarism

When introducing Korsgaard's purebred version of voluntarism in Chapter 2, I noted that her primary motivation is to explain the "authority" of reasons. After considering Enoch's "shmagency" objection, I concluded that purebred voluntarism isn't in a uniquely good position to explain the authority of reasons in general. However, it might still be that some version of voluntarism is particularly well-positioned to explain the authority of some subset of reasons. Westlund develops this line of reasoning in defense of the idea that we can sometimes create reasons by making a particular kind of commitment: a commitment to defer to others.

215 They are, as Sam Rickless once put it, the "dissent-ation" part of the dissertation.

To see Westlund's idea, start by considering situations in which it's appropriate to defer to another person's reasoning, i.e. to take the fact that someone else has decided something as a reason to do it. For example, suppose that you and I are planning to go out for dinner. That you've suggested we go to Bouna Forchetta might be a reason for me to agree, even if – other things equal – I'd prefer Juniper & Ivy. Similarly, if you're my doubles partner and you call for a ball, I should let you take it even if I think that's a mistake. For us to play together, I have to trust you. Others' decisions can also preempt or exclude our own reasoning. If I invite you to choose a restaurant and you tell me your choice, other things equal I shouldn't then revisit the question of where we ought to eat.

This isn't to say that it's always appropriate to defer; it's not. But sometimes it is, and when it is, we can ask: why should we put so much normative weight on others' say-so? Sometimes, what's going on in these cases is that there are other reasons that indirectly speak in favor of deference. Maybe you know more about the San Diego restaurant scene than I do. But in other cases, it's harder to see how these other kinds of reasons get a grip. In the tennis case, for instance, that we ought to defer to each others' calls seems to be settled in advance of our discovering the details of any particular rally, and *a fortiori* whether deferring would be most conducive to winning that particular point. Moreover, cases where deference looks appropriate needn't hinge on a previous agreement. I might decide to let you choose a restaurant without ever consulting you about who should choose.

Westlund suggests that we can explain these harder cases if we accept a voluntarist view, where (at least sometimes) we can make a consideration count as a reason by act of will.[216] If I

216 See Westlund (2013). Westlund herself frames this as an argument that we have normative powers rather than

can do this, then my decision to defer to you on some matter can explain why what you say goes, even when the ordinary grounds for deferring point another direction. Perhaps some activities, like playing cooperative sports, involve commitments to treat certain kinds of utterances as reasons. If so, voluntarism helps explain a class of otherwise puzzling reasons.[217]

Westlund's argument then looks like a local version of Korsgaard's. The challenge is to explain the authority of some reason, and voluntarism looks like a promising explanation. Chang's arguments take a different tack. Rather than looking at cases where it's intuitively clear what agents ought to do and asking what could explain that judgment, Chang's arguments develop out of looking at cases where the reasons an agent has do not seem to settle what she ought to do.

Suppose that you have a big decision to make and are reflecting on your reasons. Perhaps you are deciding whether to accept a job offer that will require you to move to a new city, far from friends and family. You might start out by tallying up considerations on each side: the job might offer better pay, fun colleagues, or a chance to develop new skills, while moving might mean spending less time with your loved ones and changes to the role you've carved out for yourself in your community.

It's plausible that in situations like these, sometimes the relevant reasons don't conclusively settle what you ought to do. Perhaps, for example, there's a tie, or your reasons for

---

one about voluntarism. Given that she seems to have the conception of normative powers advocated by Owens (2012) in mind, I think it's fair to read her as committing to a kind of voluntarism. But, she might have a more externalist-friendly conception in mind like the one suggested in Watson (2009). The contrast between Owens and Watson illustrates a more general kind of worry about arguments for hybrid voluntarism, one I'll take up in the next chapter.

217 Westlund goes on to suggest that this kind of explanation also sheds some light on the limits of these kinds of reasons, i.e. what we can and can't defer about, and what counts as an abuse of authority.

taking the job and turning it down are both sufficiently good to justify acting on either.[218] Following Chang, let's call these cases where your reasons have "run out."

Sometimes when we find ourselves in situations like these, we're content to flip a coin or otherwise resolve the matter arbitrarily. But other times solutions like these seem inadequate, like they fail to take the decision sufficiently seriously.[219] You might think, for example, that you don't want to be the kind of person who lets her relationship with her family hinge on the flip of a coin, or that you don't want to have passively drifted into one career path rather than another, even if both options are good ones. In such cases, we want not just to end up acting in a way for which we have sufficient justification, but to genuinely resolve the practical conflict by coming to see one option as better-justified than the other at the moment of decision. So, sometimes, we keep reflecting in hopes of reaching such a resolution.

Chang observes that continued reflection in cases like these is familiar, and doesn't seem inappropriate.[220] But, viewed from a certain angle, it can be hard to see why we would ever be justified in doing it. After all, as best as you can tell you have considered all the reasons that bear on the decision, and they don't yield a determinate result about what to do. If this is so, what good could further reflection on them do?

One thing that might justify continued reflection is the thought that there might be

218 I'm fudging a bit here. On Chang's hybrid voluntarist view, the details about how reasons might fail to determine what we ought to do turn out to be very important for spelling out both when we can create reasons and what consequences those reasons can have. In this section, though, I'm just trying to bring out Chang's motivations for recognizing some voluntarist reasons, so I'm going to try to keep things simple. I'll have more to say about different kinds of inconclusiveness and the details of Chang's account in the next section. I'll argue later that one virtue of my argument for hybrid voluntarism as contrasted with Chang's is that they allow us to be more ecumenical about these issues.

219 The thought that cases like these sometimes deserved continued deliberation is a starting place for an argument in Chang (2009). See also Railton (1992).

220 Chang (2009) p250

reasons we missed in our initial tally, or that we've not given them all the proper weight. For example, maybe a closer look will reveal that you didn't consider some feature that makes the job appealing, or some disappointing feature of the place it'd require moving to. But, as Chang points out, this can't be all that's going on in these cases. The mere possibility of error isn't enough to explain checking your work; I might have left the stove on this morning, but the odds aren't high enough for me go home and check.[221]

Chang suggests that we can give a more complete explanation of what's going on here if we accept that reflection can sometimes create reasons. If that's possible, then the fact that the reasons you already have don't settle what to do doesn't show that there is nothing else to think about. Further reflection could introduce new reasons, which – in combination with the ones you had prior to creating them – give you a determinate answer about what to do.[222]

For example, suppose that while you're reflecting on whether to take the job offer, the fact that there is great hiking in Utah doesn't factor into your deliberation. You're not currently much of a hiker, so that you could hike somewhere isn't the sort of thing that counts in favor of you moving there. But while you're mulling things over, it occurs to you that accepting this job would be a new adventure for you, and you don't want to be the sort of person who turns down adventures or chances to expose yourself to new things. Now that you're committed to adventuring, facts about the available hikes may seem more relevant. Those represent possible adventures too, and if you're moving for adventure maybe you'll embrace your new home and start to get into hiking.[223]

221 Chang (2009) p250-251
222 Chang (2009) p256-257
223 This is a case where reflection converts a non-reason to a reason, but one might also imagine cases where a weak reason is converted to a stronger one. Chang handles this complication by distinguishing reasons not only

In short, Chang's proposal is that continued reflection when your reasons don't settle what you ought to do only looks puzzling if we think that the only job reflection is supposed to do is help us assess the reasons that there are. Once they're assessed, there's no further work for reflection to do. If instead we think that reflection can create new reasons, then there is further work reflection can do, and so continued reflection can make sense as a way of doing that work.

I noted above that sometimes what justifies continued reflection is the possibility that there are reasons we didn't include in our initial tally. After hearing this example, though, you might wonder why that explanation isn't sufficient. Perhaps what's going on here is that you really did have a reason to live where there's great hiking all along, and it's just that this reason wasn't salient until you noticed your disposition to seek adventure, or something like that. Perhaps, this objection goes, once we vividly imagine being very sure that there are no pre-existing reasons we haven't considered, then it won't seem appropriate to continue deliberating after all. In other words, perhaps your decision is just hard, rather than unresolvable. If so, we don't need to appeal to the idea that reflection can create new reasons that can resolve previously unresolvable impasses. Instead, we might reflect just because we can't tell in advance which decisions are unresolvable and which are merely hard.

As a start toward a reply, it's worth noting that continued reflection like this, at least in cases where the decision is important to us, is a pretty familiar experience. Of course, that something is familiar isn't by itself a reason to believe that it's rational; we know what it's like to violate the instrumental principle too, but that doesn't lead us to think that we don't have reasons to take the means to our ends. Nonetheless, I think attention to this experience will help defuse

by their content but also by their source. See Chang (2009) p257-258.

the worry that – once we're very sure that all the reasons have been tallied – continued reflection will seem irrational.

What we need here is a guide to deciding, when there's a conflict between a purported principle of rationality and our familiar practices of reasoning, when the problem is with the principle and when the problem is with our practice. Consider an analogy with the St. Petersburg Paradox. A casino offers you the following bet. You will flip a fair coin until it comes up tails. If the first toss comes up heads, you'll win two dollars. For each subsequent heads, your winnings double. So, you'll get nothing if the first flip is tails, but you have a 1/2 chance of winning two dollars, a 1/4 chance of winning four dollars, a 1/8 chance chance of winning eight dollars, and so on. How much does it make sense to pay in order to play this game? Summing this series gets you an infinite expected value, so betting *any* finite amount is worth it from the perspective of expected value. Nonetheless, most people are not willing to bet their net worth on the game.

The lesson of the St. Petersburg case is not that we are all irrational for being unwilling to play bets like this.[224] Instead, the lesson is that rationality isn't only a matter of maximizing expected value. So, sometimes it does count against a purported principle of rationality if it's something we systematically fail to follow. We might wonder, then, why in this case we conclude that the problem is with the principle, while in the instrumental case we conclude that the problem is with us.

I think we'll find part of an answer if we consider some different motivations one might have for not playing in St. Petersburg. Suppose that one person is just very risk-averse; she won't play even for a dollar. “I know I could win a lot, but there's a 1/2 chance that I lose this dollar,

224 Thanks to Craig McKenzie for prompting me to think about this example.

and I just can't stand the thought," she might explain. Another person is trying to balance expected value with diminishing marginal utility: "There are some amounts where my life would change a lot if I lost them, so I won't risk those even if you offer me good odds."

The first agent here seems irrational to me, while the second doesn't. Many degrees of risk-aversion might be rational, but surely maximal risk-aversion is not. And that there is this contrast, I think, explains why we're willing to revise our conception of rationality in light of failure to maximize expected value, while we're not willing to revise it in light of failure to take the means to ends. The difference is that denying that rationality is about maximizing expected value allows us to more fine-grainedly distinguish agents, only some of whom seem to be going wrong.

One might worry that this strategy of looking for more fine-grained distinctions among ways of violating a principle will end up taking the teeth out of any rational principle it's applied to. Take the instrumental principle again. We might distinguish an agent who violates the instrumental principle because it's Tuesday from an agent who violates the instrumental principle because the means to her end are pretty unpleasant. We might then think, "wow, that second agent looks better than the first" and conclude that the instrumental principle should include an exception for unpleasant means. But this makes the instrumental principle too easy to satisfy. It can happen that the badness of the means is a reason to abandon an end, but that's the extreme case; surely sometimes it's instrumentally rational to do things we don't want to do.

This example shows that when trying to decide whether to revise our practice or a principle of rationality, we can't just ask whether revising the principle allows us to make more

fine-grained distinctions. Instead, we have to ask whether the distinctions it allows us to make are intuitively ones that make a rational difference. That you would be betting your net worth does seem like a sufficient reason not to take the offer in St. Petersburg, but that the means are unpleasant doesn't necessarily seem like a sufficient reason to abandon your ends.

Let's now return to the question of whether it can be rational to continue deliberating after you're practically certain that you have tallied all your antecedent reasons. If we apply this test – if we try revising the principle and see if we can do so in a way that preserves its intuitive appeal – I think Chang's proposal will come out on the right side of the ledger.

So, consider two ways one might go on deliberating in cases like Chang's. Suppose that one agent continues reflecting by just checking and re-checking her list of reasons, even after she's very confident that she's left no stone unturned. This agent really does, I think, seem irrational. But contrast this agent with another who tries to frame the question in new ways, asks herself what kind of person she'd like to be, imaginatively simulates options to see how at home she feels in them, etc. Here it's much less clear that the agent is being irrational. This suggests that, just as there's more to betting than maximizing expected value, there's more to making big decisions than identifying and weighing your antecedent reasons. And so, there does seem to be work for an explanation like Chang's to do. Moreover, the kinds of things this second agent is doing seem like just the sorts of things Chang's explanation would call for; they're ways of figuring out what one can commit to. So, while continuing to deliberate after reasons run out may initially seem irrational, I think the problem is with the idea that rationality requires us to stop deliberating in such cases rather than with the way we actually tend to respond to them.

In any case, Chang offers a second argument for voluntarism. In particular, she suggests that accepting the voluntarist explanation of what's happening in cases where reasons have run out has another explanatory upshot. Consider the cases where no option seems definitively best, but we're nonetheless averse to choosing arbitrarily. What we decide in these cases often has downstream effects on what else we have reason to do. Of course, this kind of thing can be true of any decision: if I decide to visit Colorado, once I get there I'll have reasons to go snowshoeing that I don't have in San Diego, just because snowshoeing is available in one place and not the other. But the decisions we make in these kinds of cases seem to have effects that are more wide-ranging than that. They seem to change the sort of considerations that are reasons for us in other decision contexts; in Chang's terms, they change our "ideal rational selves."[225]

For example, suppose that you ultimately decide not to accept the job offer, and that in the course of making this decision you commit to being a devoted family member. This will have implications for the particular choice between moving and staying put, but it will also have implications for other choices where considerations about family are relevant. For example, if you're a devoted family member, then that your niece has a Little League game is a reason for you to attend, when otherwise it might not be. Similarly, if you had instead given yourself a new reason to stay by committing to a local political project, then considerations like "there is a city council meeting on Tuesday" will be reasons for you. So, it's not just that what you decide about where to live has downstream effects, but that what commitments you take on when making that decision do as well: if you had decided by flipping a coin, that you ended up not accepting the job wouldn't itself have implications for whether you ought to attend Little League games or city

225 Chang (2009) p261-263

council meetings. These changes can even outlive the choice that gives rise to them. If you decide not to move because you commit to being a devoted family member, you'll continue to have reasons to put family first even if you later end up in a different career.

These examples suggest that there are a class of reasons we sometimes have that are hard to explain without voluntarism.[226] That's because while all of our decisions have downstream effects, and downstream effects can influence the reasons we have, some decisions seem to give rise to reasons that are not explained wholly by their downstream effects. For instance, deciding not to accept the job doesn't necessarily give you a reason to go to city council meetings, only deciding not to accept the job by committing to local politics does that. So, it looks like what's explaining these reasons you've acquired is the change in your commitments.

The examples above involve diachronic changes in your reasons; what you do at one time changes the reasons you have at a later time. This introduces the possibility of two things that might change the reasons you have: changes in your will and changes in your circumstances. In order to strip out changes in your circumstances, I contrasted two ways you might end up making the same decision, by flipping a coin and by making a commitment. If your reasons are different in the latter case but what you end up doing is the same, it looks like your commitment must have been doing some of the work. Still, there's another way to get at Chang's point about self-creation, one that focuses purely on synchronic cases.

To see how this works, let's focus narrowly on what is true about your reasons at the time you decide not to accept the job offer. Suppose first that you chose by flipping a coin. The result of the coin flip settles that practical question for you, but that's all it does. It doesn't bring about a

226 Chang (2009) p260-263

change in the reasons you had for taking or not taking the job; it was just a mechanism for making an arbitrary choice. Now suppose that you choose by committing to local politics. Now, if voluntarism is true, considerations like "by staying here, I can run for school board" count in favor of your decision to stay. So, your reasons have no longer run out. After we add the new reasons to the hopper, you are better-justified in turning down the offer than accepting it. So, if there can be synchronic changes to your reasons, voluntarism would explain how that's possible.

Why accept that there can be synchronic changes like this? Chang takes it as an intuitive datum that there are some,[227] but I think we can say more. If we're going to find an answer, it can't have to do with explaining what you end up doing. After all, we don't need voluntarism to explain how an agent picks arbitrarily. What Chang's argument seems to need is an explanation of how it could matter whether a decision is non-arbitrary at the time it is made, not just whether that decision changes what we have reason to do going forward.

We can get a hint about what this what might look like from considering other situations where some justification might be good enough for one purpose but not for another. Take, for example, my dim memory that that bank is open on Sundays. In ordinary circumstances, this might be sufficient for knowledge that the bank is open on Sundays (provided that it is indeed open on Sundays).[228] But, at least on some accounts, if we add to the case that I need to deposit a check before Monday in order to avoid a large overdraft fee, it no longer seems that I know that the bank is open on Sundays.[229] On some views, what changes here is not the quality of my justification (either way it's just my dim memory) but instead whether that justification is

227 Chang (2009) p250-251

228 If you prefer an account of knowledge with extra conditions, feel free to add them, e.g. that my belief was formed by a reliable mechanism or that it meets counterfactual safety and sensitivity conditions.

229 See Moss (2018) for discussion of this kind of case and potentially analogous cases of moral encroachment.

sufficient for knowledge in my particular circumstances. So, when I'm at risk of overdraft, we don't need to change our answer to whether I'm justified in believing that the bank is open, but we may need to change our answer to other questions, like whether I ought to rely on this belief for planning purposes, or whether it's appropriate for me to assert it.[230]

This case shows that we don't just care whether or not our beliefs are justified. For some purposes, we care whether those justifications amount to knowledge. Something similar is true about practical decisions. We don't only care whether our actions are justified, we sometimes care whether they are better-justified than the alternative actions. We can see this difference reflected in how it feels to think prospectively about carrying out a decision. In the case where you arbitrarily decide to turn down the offer, it would be reasonable to feel sheepish about it. Contrast that with a case where you have really strong reasons for turning down the offer; you can look forward with a sense of purpose. So, if the will can change the reasons we have at the time of deciding, it should be able to turn a case like the former into a case like the latter. This difference won't show up in whether or not we're justified in what we're doing, but it might show up in how it's rational to feel about it. This lends some support to Chang's intuition that we might continue deliberating in hopes of changing what we have reason to do at the time, independently of reasons we might acquire as downstream effects of our decision.

Let's step back a bit. In this section, I've considered three arguments for hybrid voluntarism: Westlund's argument that it helps to explain reasons of deference, Chang's argument

230 Note that while philosophers who think such "pragmatic encroachment" on belief is possible tend to focus on the effect of one's practical circumstances on knowledge, one needn't accept that view in order to accept the point I'm making here. Other normative statuses might change, like the assertability of my belief, or the permissibility of using it as a premise in an inference or in planning. These things might go along with changes in knowledge (e.g. if knowledge is the norm of assertion), but if they don't, my circumstances could make a difference to them without having made a difference to what I know.

that it rationalizes deliberation after reasons run out, and Chang's argument that it explains the diachronic and synchronic effects of our commitments. I've also tried to highlight some premises in these arguments one might find suspect, and to offer some novel of defenses of those premises. In the next section, I'll provide some more detail about the view Chang takes these arguments to support, and raise some questions about whether these arguments support that version of hybrid voluntarism uniquely well.

## 2. Varieties of Hybrid Voluntarism

So far I have characterized hybrid voluntarism rather loosely as the view that some reasons have an externalist source and others a voluntarist source. This leaves open questions about which reasons those are. In this section I'll sketch some ways of spelling out the details, starting with Chang's version and then suggesting some alternative possibilities that I take to be just as well-supported by the arguments above.

The arguments from Chang that I presented focus primarily on cases where an agent feels torn between two options. In such cases, hybrid voluntarism might be called on to explain both how it can be appropriate for agents to continue to mull over such decisions, and how in making such decisions agents exercise some control over deep facts about who they are.

These cases suggest something about the role that role that voluntarism should play in a hybrid voluntarist theory. If agents can create reasons, their ability to do so is in some way conditioned and constrained by the reasons that there already are. Nonetheless, there is room for disagreement about what these conditions and constraints are. In this section, I'll identify some different ways we might spell out those constraints.

Let's start with a toy model and introduce complexity from there. Recall the above case about career choice. You're offered an appealing new job, but accepting the offer would require moving away from your friends and family. The reasons you have in this case, we supposed, don't produce a determinate answer to what you ought to do; neither accepting the job nor turning it down is better-supported by your reasons than the other. A flat-footed reaction to this kind of case might be to think that voluntarist reasons stand to externalist reasons as tiebreakers. Perhaps we can create them only when the externalist reasons don't settle what we ought to do. Let's call this the Tiebreaker Model of hybrid voluntarism.

One could use arguments like Chang's in support of the Tiebreaker Model: it would explain both the appropriateness of continued deliberation in cases like your career choice and how decisions like that can be exercises of a kind of freedom that goes beyond responsiveness to reasons. Nonetheless, Chang does not accept the Tiebreaker Model. Instead, she takes her arguments to support a more wide-ranging role for voluntarist reasons. To see why, we need to say more about the ways in which your externalist reasons might fail to provide a determinate answer to what you ought to do.

Chang recognizes three ways this could happen. One possibility is that our reasons support multiple options equally well.[231] Suppose that I'm betting on the outcome of a tennis tournament, and all that matters in my choice is the probability that a player will win (I'm not, say, placing the bet in order to give myself a rooting interest, or to express loyalty to a favorite player). It might happen that two players are equally likely to win, and so my reasons for betting on one are no better than my reasons for betting on the other. So, my reasons don't offer me a

231 Chang (2009) p248-249

conclusive verdict.[232]

Another possibility is that there are two or more options that are neither better, worse, nor equally good relative to each other. Chang calls these cases where the options are "on a par." Roughly, the idea here is that value doesn't always come in determinate amounts. To borrow an example from Chang: ask yourself whether Leonardo or Michelangelo was a more creative artist. It's tempting to say that there's something wrong with the question; they are both very creative, but they are *differently* creative, or creative along different dimensions. And yet, it's not that we can't make comparative judgments about creativity. Consider Schmichelangelo, Michelangelo's less talented but stylistically similar cousin. We have no problem saying that Leonardo is more creative than him. The upshot, Chang argues, is that Michelangelo and Leonardo are not incomparable with respect to creativity; it's just that "equal," "better," and "worse," are not the only values a comparison can take.[233]

A third, related way that our reasons might fail to deliver a conclusive verdict is for them to say one option is better, but by an indeterminate margin. For instance, suppose that you're looking to buy a painting, and what you care about is buying a painting by the most creative painter you can. Your options, at first, are a Schmichelangelo and a Schmeonardo. So, let's imagine, these options are on a par. But now a Leonardo comes on the market, and given his

232 Recall that Chang doesn't think we *always* need to create reasons in cases where there's no conclusive verdict; we might just flip a coin or otherwise decide arbitrarily. Perhaps this case is one where something like that is appropriate. But in principle there's no reason there couldn't be a tie in a case that's less friendly to arbitrary tie-breaking, like if tennis gambling were very important to me.

233 One might ask whether cases like these should lead us to accept the parity relation; maybe it's just really hard to tell whether Leonardo or Michelangelo is more creative. For Chang's full argument, see her (2002) and (2012). To put my cards on the table, I'm not sure whether to accept the notion of parity. It's a central part of how Chang spells out her version of hybrid voluntarism, so I'm putting in on the table here for exegetical purposes. I'll argue below that one can accept my version without going in for it.

extreme creativity, it's now your best option. How much better of an option than the other two is it? Given that there was no determinate verdict about the comparison between the Schmichelangelo and the Schmeonardo, knowing that Leonardo is better than Schmeonardo doesn't fully answer this question. It tells you that the Leonardo painting is a better option, but it doesn't tell you that it's better by a determinate amount.[234]

Both the second and third possibilities distinguish Chang's version of hybrid voluntarism from the Tiebreaker Model. In cases of parity, we don't face a determinate tie; instead, it's indeterminate what our reasons say. In cases of indeterminate inequality, our reasons do settle what we ought to do, but they don't settle how much better one alternative is compared to another. Neither of these are ties, so the Tiebreaker Model won't count them as cases where we can create reasons. And yet, because they are cases where there is some indeterminacy in our reasons, Chang's view does.

If one is focused only on the question of what to do, it might be puzzling why cases of indeterminate inequality can call for creating reasons. But notice that the kinds of phenomena Chang uses hybrid voluntarism to explain can arise in cases of indeterminate inequality just as they can in cases of parity or equality. Suppose that you are offered a job that, professionally-speaking, is too good to turn down. In such a case, even if you are attached to your home in various ways, it might be clear that those considerations don't outweigh the professional benefits. Nonetheless, thinking about the job offer might lead you to notice that you haven't really made up your mind about what your relationship to home is. It might bug you that this question about who you are is unsettled, and so it might occasion the kind of continued mulling that cases of ties

234 See Chang (2013b) p178-179

or parity do. It's just that, in this version of the case, continuing to mull over the decision is aimed at reducing indeterminacy about the gap between the options rather than about which option is better. Moreover, making up your mind about this (say, by deciding that you're committed to family and so that moving represents a kind of loss, or that you're not and it doesn't) might play the same role in constructing your identity as decisions about ties or cases of parity do. If you decide that it is a loss, the move will mean something different to you than if you decide that moving would be no big deal.

In short, the role for voluntarist reasons suggested by Chang's arguments is more that of indeterminacy-resolver than tiebreaker. Acknowledging this puts us in a position to describe her view a bit more precisely. On Chang's version of hybrid voluntarism, the creation of reasons is subject to two constraints. The first has to do with when it's possible for us to create reasons, while the second has to do with what normative consequences the reasons we create can have. Let's take them in turn.

The constraint on when we can create reasons says that we can do so in cases of equality, parity, or indeterminate inequality. Call this the *Only Indeterminacy* constraint.[235] In cases where there is indeterminacy of any of these kinds (in cases of inequality, it's indeterminacy about what to do, while in cases of parity or indeterminate inequality, it's indeterminacy about the relationship between options), we can create reasons. These reasons then bear on our current decision, and they go into the hopper of existing reasons that might be relevant to future decisions.

235 Chang (2013b) calls this the “metaphysical” constraint on when we can create reasons, as opposed to a “normative” constraint on what consequences self-created reasons can have on what we can do. I need more specific names here, since below I'll suggest alternative constraints.

It's easiest to see how this works in cases where the reason you create has different content from ones that were already in the hopper. “There is a city council meeting on Tuesday” wouldn't be a reason for you at all absent your commitment to local politics. But Chang notes that our commitments can also change the significance of considerations that already counted as reasons. For example, recall the example about coming to count “there is great hiking in Utah” as a reason to accept a job offer that would require you to move there. In the original example, the availability of hiking didn't count one way or the other for you until you committed to pursuing new adventures. Nonetheless, we might also have imagined the case in a way where this fact did already count in favor of moving; perhaps you like hiking a little, but not so much that it's a decisive consideration for you. In a case like this, it still seems that committing to exploring new places could change this consideration's normative status. On Chang's view, we should make sense of this by individuating reasons by their source as well as their content.[236] So, once you make this commitment, there are two ways in which the consideration “there is great hiking in Utah” counts as a reason for you: first in virtue of the independent facts about the goodness of hiking and second in virtue of your commitment to new adventures.

As it happens, Chang thinks that indeterminacy about value is pervasive, and so the *Only Indeterminacy* constraint is almost always satisfied.[237] So, for her, the primary way that reasons with an externalist source condition reasons with a voluntarist source is by way of a constraint on the normative consequences that reasons we create can have. This constraint – call it *No Valence Changes* – holds that the reasons we create cannot make it the case that we have sufficient reason

236 Chang (2009) p257
237 Chang (2013b) p178

to do something that we previously had decisive reason not to do.[238] As a result, in cases of equality or parity, our created reasons can make a difference to what we ought to do in our current decision-context (they might make it rational to go for one career rather than the other), but in cases of indeterminate inequality they can only make a difference to what we ought to do in other decision-contexts. For example, suppose that we all have better reason to have cream cheese on a sesame bagel for breakfast than mashed avocado on multi-grain toast, but it's indeterminate how much better bagels are. On Chang's view, by adopting the practical identity of a Californian, part of which – let's stipulate – is enthusiasm for avocado toast, I can narrow the gap between them. This won't change what I have most reason to do when offered a fresh bagel, but it can matter going forward when I encounter related choice situations. If avocado toast is almost as good, I should choose it over a stale bagel. If it pales it comparison, the stale bagel is the way to go.

So, putting things together, the *Only Indeterminacy* constraint rules out creating reasons in cases where our antecedent reasons produce a determinate verdict by a determinate margin, and the *No Valence Changes* constraint requires that when we create reasons, the difference they can make depends on the kind of indeterminacy they were created to resolve. When the verdict is what's indeterminate, self-created reasons can change it; when only the margin is indeterminate, self-created reasons can only change that.

One result of this pair of constraints is that Chang's hybrid voluntarism avoids the extensional difficulties that Korsgaard's purebred voluntarism encountered. Recall the Mafioso whose commitment to a life of crime requires taking considerations about what will harm her

238 Chang (2013b) p179

rivals to be reasons to bring about those harms. Because Chang recognizes externalist reasons, she can say that the Mafioso has decisive reason not to do these harmful things, and so the Mafioso's commitment can't make it the case that she ought to do those things. That would violate *No Valence Changes*. But, Chang notes, this view has a deeper motivation. Specifically, it makes our control over our normative situation analogous to our control over our physical situation; we can move around freely, but we're constrained by independent features of the physical world, like walls.[239] So, how much freedom we have to create reasons is going to depend on what other reasons we have, just as how much freedom we have to move around depends on what other physical stuff there is.

This analogy has intuitive appeal, at least for me. But I don't think it speaks unequivocally in favor of Chang's view. In particular, I think we can respect the appeal of this analogy, and the role Chang's constraints play in avoiding extensional problems like the Mafioso, while weakening both *Only Indeterminacy* and *No Valence Changes*. In the remainder of this section I'll explain how.

Let's start with *No Valence Changes*. The effect of this constraint is to draw a sharp boundary between cases of equality and parity, on the one hand, and indeterminate inequality, on the other. In the former we can create reasons that change what we ought to do, while in the latter we can create reasons but not ones that would have this effect. Put in terms of the physical space analogy, *No Valence Changes* represents verdicts about what to do as immovable walls; we can push on them, but they won't fall over. Notice, however, that there are other ways we might interpret the physical space analogy. Not all physical walls are insurmountable obstacles; some

239 Chang (2013b) p179

we can knock down or jump over. So, even if we think of verdicts given by our antecedent reasons as like physical walls, the physical space analogy doesn't force us to think of them as absolute constraints. We might instead look for a view where we can knock over flimsy walls but not sturdy ones.

The same seems to me to be true about potential extensional objections to hybrid voluntarism. Consider two cases of indeterminate inequality. Given the differences between the benefits and burdens of the two careers, being a philosopher might be better for me than being a lawyer, but indeterminately so. Being a philosopher might also be better for me than being a Mafioso, but indeterminately so. Still, it's not clear to me that these cases can ground the same kind of extensional objection to hybrid voluntarism. Admitting that I could create reasons that tip the scales in favor of being a Mafioso would be very counterintuitive, but admitting that I could create reasons that tip the scales in favor of being a lawyer doesn't seem so bad. Like the question about how to interpret the physical space analogy, the question about whether voluntarist reasons could make a difference to what I ought to do in some cases of indeterminate inequality seems like an open question, or at least not one that's settled by thinking about the kinds of cases that make trouble for purebred voluntarism.

Of course, if there were no constraints on the normative consequences of self-created reasons, hybrid voluntarism would have the same extensional difficulties that purebred voluntarism does. So, given that the physical space analogy leaves open the choice between *No Valence Changes* and a weaker constraint, the question of whether there are plausible alternatives to *No Valence Changes* boils down to the question of whether a weaker constraint could also

avoid these extensional difficulties.

I think there's one that can. The key to spelling it out is to notice that indeterminate inequalities still admit of some kinds of comparison. Consider a creativity comparison between Leonardo and Schmichelangelo. Leonardo is better, but by an indeterminate amount. Now consider a creativity comparison between Schmichelangelo and a really terrible painter. Suppose that this comparison also yields an indeterminate inequality: Schmichelangelo is better, but not by a determinate margin. Taking these comparisons together, it looks like we should say that – though the margins are indeterminate – the creativity gap between Leonardo and the terrible painter is bigger than the creativity gap between Schmichelangelo and the terrible painter.

Admittedly, it's odd to think of indeterminate comparisons as having different magnitudes. But not *that* odd, I think. To say that some value isn't totally determinate is not to say that nothing can be said about its contours. We might, for example, be able to say something about the different dimensions of our creativity judgments, without being able to give an exhaustive list of creativity-relevant features or the ways they might interact.

We can use this observation about indeterminate inequalities coming in different sizes to make space for an alternative constraint on the normative consequences of self-created reasons. Given that indeterminate inequalities can have different magnitudes, we have grounds to treat them differently. Consider again my choice between being a philosopher and a lawyer, versus my choice between being a philosopher and a Mafioso. They're both cases of indeterminate inequality, but given that the gap between philosopher and lawyer is less significant than the gap between philosopher and Mafioso, we can admit this and still have grounds for thinking that

voluntarist reasons could tip the scales in one case but not the other.

Can we make this more precise? Given that we're talking about indeterminacy, it's hard. Still, here's a rough schema: we might say that the weight of a new voluntarist reason to perform a given action is inversely proportional to the aggregate weight of one's existing reasons against performing that action. In cases like the choice between being a philosopher and being a Mafioso, the reasons against being a Mafioso are very strong, and so the reasons I can create in favor of being a Mafioso are weak. In cases like the choice between being a philosopher and being a lawyer, the reasons against being a lawyer are comparatively weak, and so the reasons I can create in favor of being a lawyer are comparatively strong. This leaves open the possibility that, in cases where the reasons against one option are sufficiently weak – but still strong enough to, absent will-based reasons, yield a verdict about what to do – will-based reasons could be strong enough to change that verdict. We might call this approach *No Dramatic Valence Changes*, since it prohibits valence changes when indeterminate inequalities are large (compare: when the walls are sturdy) but allows them when they're not (compare: when the walls are flimsy).[240]

*No Dramatic Valence Changes* spells out a constraint on voluntarist reasons in a way that's independent of the content of the reasons they compete with. Strong reasons not to be a

240 This way of spelling out *No Dramatic Valence Changes* appeals to the idea of a reason's weight, but "weight" is ambiguous here. Suppose I'm choosing between being a Mafioso and a white collar criminal. In one sense, there are strong reasons against doing both of these things; both kinds of crime are bad. In another sense, the reasons I have not to be a white collar criminal are not very strong here, not compared to my reasons not to be a Mafioso. What's going on here is that reasons can be strong *in a decision-context* if they make a big difference to settling what you ought to do in that context, but they can also be strong *across decision-contexts* if they make a big difference to settling what you ought to do in related contexts. My reasons not to be a white collar criminal are strong in the latter sense (most of the time, they're decisive) but not in the former (they're not decisive against my reasons not to be a Mafioso). In spelling out *No Dramatic Valence Changes*, I have weight across decision-contexts in mind. Thanks to Ruth Chang for pressing me on this.

lawyer constrain the creation of reasons in the same way that strong reasons not to be a Mafioso would. But once one recognizes the possibility of scalar constraints, one might choose to articulate them in a more fine-grained way. Perhaps, for example, something like *No Valence Changes* is true for conflicts between voluntarist reasons and moral reasons, while something like *No Dramatic Valence Changes* is true for conflicts between voluntarist reasons and prudential reasons. To push the physical space analogy a bit further, we might think of moral reasons as load-bearing walls, while other reasons with an externalist source are more open to renovation.

So far we've focused on the "walls" part of the physical space analogy. What about how hard you push? One plausible option is to understand pushing in terms of the centrality of a commitment to the practical identity of the agent-making it. One of my hobbies is brewing beer, another is following baseball. In some ways, baseball is more central to my practical identity. It's resonant with other projects (e.g. papers in the philosophy of sports), it's a part of more of my friendships (e.g. watching games together), and I devote more time and intellectual energy to it (e.g. reading sabermetrics). Plausibly, then, reasons I have in virtue of a commitment to following baseball are more apt to outweigh reasons with an externalist source than reasons I have in virtue of a commitment to home-brewing are. So, just as we can finesse our views about the relationship between externalist and voluntarist reasons by appeal to the content of the externalist reasons, so too *No Dramatic Valence Changes* is compatible with finessing things by appeal to the agent's relationship to the voluntarist reasons in question.

Note that *No Dramatic Valence Changes* does not amount to saying that voluntarist

reasons can only change what we ought to do in cases of parity. Relatively small indeterminate inequalities are still inequalities, just as flimsy walls are still walls. So, a version of hybrid voluntarism that accepted this constraint would permit voluntarist reasons to have normative consequences that Chang's version of hybrid voluntarism would not.

So, I think, it's possible to reject *No Valence Changes* but still end up with a version of hybrid voluntarism that (a) avoids the kinds of extensional problems that cause trouble for purebred voluntarism and (b) respects the physical space analogy. Of course, I haven't argued that we should accept this alternative version, just that extensional considerations and the physical space analogy don't rule it out. To figure out whether which version of hybrid voluntarism is most promising, we'd have to argue on other grounds. I'll suggest some ways of starting that argument in the next chapter.

Let's turn our attention to Chang's other constraint, *Only Indeterminacy.* Recall that this constraint says that we can create reasons only in situations where reasons run out, i.e. when there are ties, parity, or indeterminate inequality. As a result, we cannot create reasons when our antecedent reasons deliver a completely determinate verdict. For example, suppose that I am completely indifferent about which shoe I ought to tie first; I don't have any reasons that bear on this one way or the other. But then, you offer me $5 to tie my left shoe first, and someone else offers me $100 to tie my right shoe first. The only reasons in play here are the financial ones, and $100 is more than $5 by a determinate margin. Given that I don't care about anything else in this situation, $100 is therefore better than $5 by a determinate margin, too. So, in a case like this *Only Indeterminacy* applies and it's not open to me to create any new reasons that might bear on

this choice.

This might seem like a contrived example, but that's intentional. Recall that Chang thinks that parity and indeterminate inequality are pervasive, so in more realistic choice situations *Only Indeterminacy* is almost always going to be satisfied.[241] Still, it's worth acknowledging that the ideas of parity and indeterminate inequality are controversial, and I haven't done much to defend them here. So, for all I've said, you might think that either (a) cases of parity and indeterminate inequality are really epistemic rather than metaphysical phenomena, or (b) that cases of parity and indeterminate inequality are rare, arising only with respect to a particular subset of complex values like creativity. If either of those things is true, then *Only Indeterminacy* gets more constraining.

In any case, I think that *Only Indeterminacy* can be relaxed, for much the same reasons as with *No Valence Changes*. To see how such a view might work, note that even setting aside cases of parity, *Only Indeterminacy* is going to be satisfied in cases of ties. For example, if in the shoelace case you offer me $5 to tie my left shoe first while someone else offers me $5 to tie my right shoe first, this will still be a case of reasons running out.

In exploring ways to relax *No Valence Changes*, I noted that we can respect extensional adequacy and the physical space analogy even if agents can create reasons when they face indeterminate inequalities between options, so long as those inequalities are comparatively small. So far as I can tell, the same kind of point applies to *Only Indeterminacy*. Some determinate inequalities are bigger than others, and so some pairs of options are closer to equally good than others. This gives us grounds for treating different cases of inequality differently. Just as we

241 Chang (2013b) p178

might think it's more plausible for will-based reasons to tip the scales when I'm choosing between being a philosopher and a lawyer versus choosing between being a philosopher and a Mafioso, we might think it's more plausible that I could create will-based reasons when I'm deciding between a $5.00 right-shoe offer and a $5.02 left-shoe offer versus a $5.00 right-shoe offer and a $100 left-shoe offer. In short, one might hold that we can create reasons only when our options are sufficiently close to equally good. We might call this strategy *Indeterminacy or Proximity*, since it allows for the creation of will-based reasons when one's options are indeterminate in some way, equal, or close to equal.

In short, I think, one can respect the physical space analogy and extensional adequacy while being more permissive than Chang about both when we can create reasons and what effects those reasons can have. We can do the former by trading *Only Indeterminacy* for *Indeterminacy or Proximity* and we can do the latter by trading *No Valence Changes* for *No Dramatic Valence Changes*. These trades are motivated by the same kinds of considerations – the idea that not all inequalities are equally decisive when it comes to what we ought to do – but in principle they're separable. For instance, if you find it more plausible that the we can create reasons in cases of indeterminacy than in cases of determinacy (be they determinate ties or determinate inequalities), you could hold on to *Only Indeterminacy* as a condition on when we can create reasons but still accept *No Dramatic Valence Changes* as an account of what normative consequences they can have.

One interesting feature of accepting *Indeterminacy or Proximity* is that it decouples the plausibility of hybrid voluntarism from the plausibility of parity and indeterminate inequality. To

a parity-skeptic, apparent cases of parity might instead look like cases where it's hard to tell which of two options is better. We might still, though, be able to tell that two options are close to equal. Either way, then, *Indeterminacy or Proximity* will count these as cases where it's possible to will a reason. In the one case it'll be because there's indeterminacy, and in the other it'll be because there's proximity. So, even a parity-skeptic can accept a version of hybrid voluntarism that goes beyond the Tiebreaker Model, and so this parity-skeptic can get some of the explanatory payoffs of hybrid voluntarism I'll explore in the next chapter.

So far, I've just been arguing that Chang's conditions *can* be relaxed; I've left alone the question of whether they *should* be. To put my cards on the table: my preferred version of hybrid voluntarism relaxes both conditions. Nonetheless, my primary goal here is to bring out the plausibility of hybrid voluntarism as a general family of views. So, in the next chapter I'll develop some arguments that I think lend plausibility to both my view and to Chang's. Once those arguments are on the table, I'll return to the question of whether they support one of our views more strongly than the other, and what other questions we might need to investigate to decide between them.

# Chapter 6

## New Arguments for Hybrid Voluntarism

In Chapter 5, I sketched some existing arguments for hybrid voluntarism: it helps to explain why it can be rational to continue deliberating even after you have assessed your reasons, it helps explain how it can be rational to defer to others even when what they tell us to do is suboptimal, and it seems like a natural explanation for the effects our commitments have on what we have reason to do. In Chapter 2, we also saw that at least one motivation for purebred voluntarist views – that they help explain the phenomenology of deliberation, especially the way that acting on your reasons feels closely tied to being yourself – can be marshaled in favor of hybrid voluntarism as well. In the rest of this chapter, I want to suggest some other lines of argument in favor of hybrid voluntarism.

My argumentative strategy here proceeds in two stages. First, I'm going to argue that the will is normatively significant; in other words, changes in an agent's will can make a difference to what that agent ought to do. On its own, this constitutes an argument against several of the views we've considered so far. It's inconsistent with internalism, since it allows that something other than a change in desires (or facts about about what would promote the satisfaction of them) could produce a change in reasons, and it's inconsistent with flat-footed versions of externalism, since it allows that there's something relevant to our reasons other than the objects of our choices. Nonetheless, the claim that the will is normatively significant is consistent with some more sophisticated versions of externalism, according to which the act of choice can be valuable in addition to the objects of choice. So, in the second stage, I'll argue for the further claim that: if

the will is normatively significant, the most plausible explanation for this is that the will is metanormatively significant too. If this is correct, hybrid voluntarists give us a better explanation of the will's normative significance than do these sophisticated externalists.

In earlier chapters, my arguments depended on pointing to cases that particular views have a hard time accommodating. My argument here will take a different tack. I won't be focusing on cases that I think externalists get wrong; given their non-reductive ambitions, externalists have an easy time getting cases right. Instead, I'll be arguing that hybrid voluntarists explain the phenomena in a more satisfying way. So, my argument here will depend on some methodological claims about what kind of explanation would be best. In any case, before we can compare explanations, we should identify the phenomenon to be explained.

## 1. First Pass: The Normative Significance of the Will

The first thing on the agenda is to argue that the will is normatively significant. I'll develop two arguments for that here. The first of these arguments is an extension of Chang's suggestion that we individuate reasons based on their source as well as their content. Chang introduces this idea in order to make sense of what happens when a commitment we make changes the significance of a consideration that was already a reason for us. I think that attention to this possibility can also help explain what's going on in some kinds of disagreements about reasons.

Consider the sorts of things we often say when asked what's appealing about a career in philosophy: it's rewarding to participate constructively in an intellectual tradition, it gives us the opportunity to talk to really smart people all day, it's fun to help students become clearer

thinkers, etc. Plausibly, these really are good things about doing philosophy, and they would be good for anyone with the relevant aptitudes and work habits. Nonetheless, when one finds oneself rehearsing them in certain contexts (say, talking about graduate school with your lawyer in-laws or parents' skeptical friends), it is easy to get the feeling that they are not as convincing as one might hope. What seems to be going on here is that one's interlocutors are prepared to grant that these things are true about working in philosophy (you really do get to help students, they'll admit) and that they really are reasons (that really is a good thing to do) but not that such considerations could be decisive for you (how could that outweigh the risky job market, they'll wonder).

If we grant that the will is normatively significant, we get a nice explanation of how two people can seem to recognize the same consideration as a reason but nonetheless feel the force of it so differently. What's going on, we can say, is that the consideration "philosophers get to help their students think more clearly" does not have the same normative status for both parties in this conversation, despite the fact that it is a reason for both. For me, it's both an independent good and a central part of my practical outlook, the kind of project I've organized the last several years of my life around (reading about pedagogy, observing and talking with peers, experimenting in class, and so on). So, this kind of consideration is really two reasons for me: a (possibly not so weighty) one I have as a result of my circumstances, and a (very weighty) one I have as a result of my will. If this is what's going on, it makes sense that when talking to people with sufficiently different commitments, they will recognize these as reasons but not feel their force the same way I do. They'll recognize them because that kind of consideration really is a reason for them, but

not feel the same force because most of the force they have for me comes from a commitment my interlocutors don't share.

The case of a career one feels called to is a salient example of this for philosophers, but I think this kind of disagreement about reasons is pervasive. Many hobbies, for example, are such that I can see why they are appealing but don't myself feel grabbed by them. If the will is normatively significant, this is a predictable upshot of the different ways in which we exercise our wills.

One might think that we don't need to think that the will is normatively significant in order to explain these kinds of disagreements. Perhaps we can instead get by with the distinction between agent-neutral and agent-relative reasons. Maybe working with very smart people is good for anyone, and particularly good for philosophers in light of some particular role we occupy (maybe as people who value collaboration, or something along those lines). This might be what's going on in the case above, but I think we can imagine cases where the same phenomenon occurs (i.e. where the same consideration looks like it amounts to two reasons), but where agent-relativity doesn't obviously help.

The trick to imagining that kind of example is to start already squarely in the realm of the agent-relative. Consider a case where the independent normative facts don't provide reasons for *anyone* to do something, but instead provide reasons for anyone in a certain role to do something. Perhaps anyone who finds herself in the role of "member of the philosophy department" has agent-relative reasons to do things that support the department community: attending talks, commenting on colleagues' drafts, hosting social events, and so on. It seems to me that these

kinds of agent-relative reasons are subject to the same kind of phenomena as the agent-neutral one above. When Noel and I were graduate representatives, that role gave us some reason to throw a party to welcome the incoming first-year class, but not a decisive reason given the expense of hosting a party. Nonetheless, for some people, these kinds of reasons seem to count twice, and they seem to do so exactly when those people are committed to certain projects, e.g. as promoters of an inclusive workplace or as friends to the incoming class. So, even when we're already squarely in the realm of the agent relative – no non-department-member has reasons to do these things – there is still something puzzling about individuating reasons solely by their content. But if we think that the will is normatively significant, we can appeal to both the agent-relative reasons that apply one has as a result of one's role and the agent-relative reasons one has as a result of one's will. So, it's easier to explain the double-counting that seems to be going on in these cases.

Still, one might grant the point that some agent-relative reasons are susceptible to this same sort of double-counting phenomena, but insist that in those cases there are really just more will-independent reasons at play. For example, “hosting this party would help the first-years meet people” might be an (agent-relative) reason for Noel and me both as department representatives and as promoters of an inclusive workplace. If so, we don't need to appeal to the will to explain why the same consideration can count as two reasons for one person and only one for another; we can just appeal to other role-based agent-relative reasons.

This strategy strikes me as ad hoc. It's not a coincidence, I think, that the examples that generate these double-counting situations involve the kind of commitments or personal projects

that have tended to motivate voluntarists. One might ask, for example, why one person has the role such that "hosting this party would help the first-years meet people" is especially salient for them. The natural answer is that they've committed to that role. So, it seems, to capture the generality of this phenomenon we should think that the will is normatively significant, at least in some range of cases.[242]

Another argument for the will's normative significance concerns normative motivation. Recall the questions about normative motivation we considered in Chapter 4. There, we saw that part of externalists' strategy for explaining normative motivation is to make a constitutive claim about agency: part of what it is to be an agent is to be motivated by your judgments about reasons. While this claim is very plausible, it could be more satisfying as an explanation if we could say more about why this constitutive relation holds. Otherwise, we have an explanation of the mechanisms by which agents are motivated to do what they judge they ought, but no bigger-picture explanation of why being a rational agent involves having those mechanisms. Recognizing the will's normative significance provides the resources for an explanation of just this sort.

Consider the following two scenarios. In both, you are part of a committee that will determine who to hire for a new position. Once a decision is made, as a committee member you will have various obligations related to executing the decision, e.g. recruiting the candidate if she has other offers, helping her get settled in your workplace, etc. In the first scenario, you have felt ignored during most of the committee's deliberations: your concerns about candidates are

242 Recall that so far this is just a normative claim. For this section, it could be true either because there is something mind-independently valuable about exercising one's agency (i.e. an externalist story) or because the will can sometimes make a consideration a reason (i.e. a hybrid voluntarist story). I'll argue for the latter over the former in the next section.

minimized, your suggestions about how to ensure a fair decision procedure aren't discussed, you don't get to weigh in on the desiderata for a candidate, and so on. In the second scenario, the opposite is true. Your objections are heard and engaged, you were consulted about procedure and the goals of the committee, and so on. Very plausibly, you'll be more motivated to carry out your committee member duties in the second scenario than in the first.

The contrast between these cases illustrates a more general claim: we tend to be motivated to act in accordance with a normative standard when we have some say in what the standard is. Importantly, this seems true independently of whether or not the standard gets set in the way we prefer; you could feel heard in the hiring committee process even if the committee ultimately decides on a candidate of whom you're not a fan. Though getting your way also matters, having a voice in a process can be motivating without it.

It also seems to matter that it be you who has the voice in the process, rather than just that your ideas or interests are represented. Suppose that you have a colleague who agrees with you on everything hiring-related, but that during the meetings it's always your colleague's arguments that get uptake in discussion. This situation seems worse, motivationally speaking, than the one where your arguments get uptake too.[243] Here again, motivation to act on the committee's decision tracks your involvement in it.

Here's the point of this detour into armchair psychology: if the will is normatively significant, then we all have a voice in what we have reason to do. So, deliberating about what you have reason to do is more like being on the functional committee than the marginalizing one.

243 Of course, the effects on your motivation aren't the only bad thing going on in this example. Depending on the details, it's plausible that your colleagues are committing a kind of epistemic injustice against you, or at least a violation of collegiality. But the feature of the example relevant to my argument here is the motivational one.

And if so, we should expect you to be motivated to do what you have reason to do, even in cases where what you have reason to do is not what you want to do.

To forestall a possible misunderstanding: my claim here is not that reasons we have independently of our wills aren't necessarily motivating, but that reasons related to our wills are. That wouldn't answer our question about normative motivation, since not all cases of normative motivation involve reasons we have as a result of our wills. Instead, my claim is that recognizing the normative significance of the will helps explain why all of our reasons are themselves motivating, including ones we have independently of our wills. They're motivating because of their role in a process in which we have a say, even if we do not have a say directly over them.

One might worry, though, that this explanation is too hasty. Even if it's true that having a say in some process can have motivational effects, it might not always. For example, consider an authoritarian government, where almost every detail of life is legislated according to a despotic leader's whims. Suppose that you find out that the despotic leader has left it up to you whether you go grocery shopping before or after taking the cat to the vet. Though this gives you (marginally) more say in how your day will go, it's plausible that it would not have the effect of increasing your motivation to comply with the despot's whims.

It seems to me that the relevant difference between this case and the functional committee case is a matter of scope. In the committee case, both (a) you have a say in more things, and (b) the things that you have a say in are more important. The despotic leader wouldn't be quite so despotic if you were given authority over more of your day, and if the decisions you could make were ones you cared about. So, assessing whether this objection goes through is a matter of

assessing the scope of the claim that the will is normatively significant. If it's wide, and if the circumstances in which the will is normatively significant include choices that are important to you, then the committee analogy holds.

Thankfully, I think the views we've been considering here do well on both counts. While it's in part an empirical question how often our reasons run out (or come close to it), a normal life will contain many such circumstances. Someone with a terribly restricted range of options might not, but for the rest of us there are many kinds of lives worth living and, on a smaller scale, many ways of arranging commitments within those lives. So, it's very plausible that if the will is normatively significant, then this significance will make a difference in a wide range of cases and in decision-contexts that we care about. If that's right, the analogy with the despotic leader is no threat to this explanation of normative motivation.

## 2. Second Pass: The Metanormative Significance of the Will

So far, I've tried to identify two explanatory payoffs of including thinking that the will is normatively significant. Specifically, I suggested that if it is, we get new resources for explaining how agents are motivated by their normative beliefs and what's going on in certain cases of disagreement about the force of reasons.

Still, one might agree with me that recognizing some link between the will and our reasons helps explain these phenomena, but nonetheless disagree about whether this gives us any reason to accept hybrid voluntarism. Instead, one might think, the best way to account for the significance of the will is solely at the *normative* rather than *metanormative* level. Perhaps, schematically, there are externalist facts of the following form: in the appropriate circumstances,

if we will such-and-such we get reasons thus-and-so. On a view like this, changes in our will produce changes in our reasons not by being the source of those reasons, but by triggering the application of some other, already normative principle. A view like that looks like it can say everything I've wanted to say so far.

To get a better grip on this contrast, it may help to compare it to an analogous one. According to internalism, what makes some consideration a reason is its relation to an agent's desires. So, for internalists, desires determine our reasons by being the thing in virtue of which considerations come to be reasons in the first place. Externalists don't accept this explanatory link between desires and reasons, but this needn't stop them from thinking that desires are sometimes normatively significant. They might think that, in the appropriate circumstances and for the appropriate desires, one of the normative facts is that you have a reason to do what you want. So, any given desire I have might be normatively significant, but not all on its own – only with the help of the relevant normative facts. Such an externalist would be denying the metanormative significance of desire, but nonetheless granting desires some normative significance.

The interlocutor I want to engage with here takes just that stance toward voluntarism, saying, “You may have shown that recognizing a connection between the will and reasons helps us to explain some things, but why think that the connection is a metanormative rather than normative one?” Call this objector the Externalist Accommodationist. In this section I'll try to provide an answer.

Ultimately, I don't think there's anything incoherent about the Externalist

Accommodationist strategy, and – given the two arguments in the previous section – I think it's better than a flat-footed externalist one. But I do think that hybrid voluntarism offers a more satisfying explanation of the will's normative significance. So, in this section I'll try to bring out what's more satisfying about it. In other words, I'll argue that if the will is normatively significant, the most plausible explanation for this is that hybrid voluntarism is true. If so, the above arguments for the normative significance of the will can also serve as arguments for hybrid voluntarism.

Still, the details of this comparison may depend on exactly what the Externalist Accommodationist wants to say. So, we need to start by distinguishing different ways the accommodation might go. Broadly speaking, I think there are two options. The difference is in where in their first-order normative theory each Accommodationist locates the will. One possibility is that while the will cannot make a consideration a reason when it wouldn't be one otherwise, it can change the weight of a consideration that is already a reason. Dale Dorsey defends a view like this.[244] Another possibility is that while the will can't make a consideration a reason when it wouldn't be otherwise, agents generally have reasons to stick with the choices they make. So, even if an agent lacked justification for a choice at the time of making it, she might have justification for continuing along her chosen path. Though he stops short of explicitly endorsing it, Brink provides a sympathetic sketch of how a view like this might go.[245]

Let's take Dorsey's proposal first. On his view, the externalist facts settle what reasons an

244 See Dorsey (2016).

245 See Brink (2008). Note that the suggestion here isn't about *structural* rationality; it's not that after having made a choice, internal coherence subjectively counts in favor of sticking to it. Instead, it's that there's some substantive principle that speaks in favor of doing what you've chosen, like substantive principles of prudence speak in favor of looking out for your future well-being. So, on this view, the reasons you have to do what you've chosen are genuinely objective and justifying.

agent has, but an agent's act of will can change the weight of those reasons. Suppose, once again, that you are choosing whether to accept a new job that will require moving far from home. It's not up to you, Dorsey thinks, what counts in favor of or against each option (say, that the new challenge counts in favor of moving but your family ties count against), but it is up to you how much these considerations count in the direction they do.[246] In other words, while the only reasons that exist are ones with an externalist source, acts of will can amplify or mute those reasons. So, acts of will can have consequences for what we ought to do (by, say, amplifying a reason that otherwise wouldn't settle that question) but acts of will don't have any effect on what reasons there are.

In arguing for this view, Dorsey claims that in all cases where voluntarist explanations of reasons look plausible, what's actually going on is that the consideration already was a reason, just a weak one.[247] For instance, Dorsey thinks, it sounds plausible to say that by committing to a relationship with someone you create reasons to promote their well-being (say, by donating your kidney when they're sick) because, really, the fact that some action would promote someone's well-being was already a reason to do it. It's just that, when you're not committed to that person, that reason isn't salient. After all, if you could get someone the kidney they need by costlessly pressing a button, their need would count in favor of doing so.

If Dorsey is right about this, it would lend some strong support to his kind of Externalist

246 Recall that Chang distinguishes reasons by their source as well as their content. If one were to insist on individuating them solely by their content, then her view might look like it substantially overlaps with Dorsey's. What she describes as creating a reason with the same content as a pre-existing reason, Dorsey might describe as increasing the weight of this reason. Does this overlap make Dorsey a voluntarist? I don't think so. Dorsey opts for his account of the weight of reasons precisely because he finds voluntarism about the existence of reasons implausible.

247 For his defense of this idea, see section two of Dorsey (2016).

Accommodation. If all the cases where voluntarism looks plausible can be recast as cases of changing the weight of a reason rather than creating a new one, then any argument drawing from those cases supports his view just as much as, say, Chang's. Nonetheless, I am skeptical that all cases where voluntarism looks plausible can be recast in this way. To continue with the kidney case, suppose that it's Cecil whose commitment to Carlos looks like it creates a reason for Cecil to give Carlos his kidney. I agree with Dorsey that, in general, the fact that Carlos needs a kidney is a reason for anyone to help him. But, I don't think a weightier version of *that* reason is the one that Cecil acquires when he commits. After all, that reason's force is totally independent of who needs the kidney. We could describe that reason to Cecil by saying that "someone needs a kidney" without any change in its normative force. In contrast, the reason Cecil acquires when he commits to Carlos is an agent-relative one; it matters that it's *Carlos* who needs the kidney. So, the change here isn't just a change in weight; it's also a change in scope. I share Cecil's reason to give anyone a kidney, but not his reason to give one to Carlos.

Moreover, Cecil's original, defeasible reason to give a kidney to anyone who needs one doesn't go away after he commits to Carlos. Were Cecil to face a choice between donating a kidney to Carlos and to a stranger, there would still be something to be said in favor of helping the stranger (just not as much as can be said in favor of helping Carlos). So, I think, it's implausible that there's really just one reason here. We capture the intuitive data more neatly if we say that there are two reasons on the scene, one to help anyone who needs a kidney, and one to help Carlos when he needs a kidney.

So, I'm not sure that Dorsey's proposal really gives us a natural description of the cases

it's intended to. Moreover, I think that Dorsey's view leads him in to some extensional trouble. The problem is that, for many actions that intuitively we have decisive reason not to do, there is nonetheless some reason in favor of doing them. For instance, suppose I am deliberating about whether to become a Mafioso. There are very strong reasons, both moral and prudential, not to do that. But, there nonetheless might be *something* to be said for it; maybe it will be thrilling, profitable, or just give me a clearer direction in life. Since Dorsey's view allows agents to change the weight of their reasons, nothing seems to rule out my making a pro-Mafioso reason arbitrarily weighty, such that – contrary to our commonsense judgment about the case – I ought to become a Mafioso after all. So, it looks like Dorsey's version of Externalist Accommodation leads to the same trouble that Korsgaard's version of purebred voluntarism does.

Of course, this feature is idiosyncratic to Dorsey's view. In the same way that hybrid voluntarists can put constraints on the creation of reasons in order to avoid extensional trouble, so too could someone adapt Dorsey's view and put constraints on the re-weighting of reasons in order to avoid extensional trouble. One might, for example, say that we can only change the weight of reasons when, using their original weights, they run out. Making this change, though, brings us much closer to Brink's version of Externalist Accommodation. So, let's turn our attention to that.

Unlike Dorsey, Brink gives a unified account of reasons' existence and weight, where both are given by the externalist facts. His proposal, instead, is that if we want to account for the normative significance of the will, externalists should think that the objects of choice are not the only things that are valuable; the exercise of the capacity to choose is valuable as well.[248] If so,

248 Brink (2008) p40-45.

that we have chosen something can itself be a reason to continue to pursue it. Brink canvasses a few different ways we might understand how these reasons make a difference to what one ought to do, but each is designed to avoid the kinds of extensional difficulties faced by Korsgaard and Dorsey.

For Brink, this proposal emerges out of a larger project of normative perfectionism.[249] On his view, what an agent ultimately has reason to do is to develop the capacities that make her an agent, capacities like recognizing and being motivated by reasons, carrying out temporally extended projects, engaging in shared deliberation with others, and developing skills. Given the central importance this view places on exercises of agency, Brink suggests, it makes sense that particular exercises of agency should also have some normative significance. So, that you have chosen something – whether or not you had most reason to choose it at the time – can count in favor of sticking to that choice later. This is, one might think, a way of taking one's own exercises of agency seriously.

Nonetheless, an externalist need not accept Brink's particular perfectionist account of the personal good in order to pursue this kind of strategy. One might think, for example, that self-government or integrity is fundamentally valuable, and appeal to this value in explaining the normative significance of the will. For instance, it might be that we typically have reasons to stick with our choices because doing so is required by the norms of planning and temporally extended agency.[250] Certain kinds of temporally-extended projects are only possible when we treat certain decisions as fixed points, rather than revisiting them at each opportunity. For

249 See Brink (2008) p31-36 for a sketch of this view and some of its historical antecedents.
250 See Bratman (2018) for a big-picture overview of this approach.

instance, if I'm going to pay the electric bill I'm going to need to get to campus on time to teach, and if I'm going to get to campus on time I'm going to need to finish my coffee. But, I won't finish my coffee if I sip it while mulling over whether or not I really ought to pay the electric bill, or whether teaching is the right way to do that. To execute the bigger-picture plan, I need to insulate some parts of it from revision. If so, perhaps the reasons we acquire when we commit get their normativity from the value of this kind of planning behavior.

Alternatively, one might think about choices about how to act on one's own values analogously to the way Ronald Dworkin suggests that judges should think about interpreting the law. On Dworkin's view, judicial interpretation should be guided by a conception of the law's *integrity*. So, when there is a question about how the law applies to a new case, judges need to figure out which interpretation of past decisions presents them in the best light, as coherently expressing a set of attractive values.[251] As a result, judges are constrained both by the values and by the precedent. So, a judge should not interpret the law so as to make it unfair, nor should a judge interpret the law so as to be incoherent with past decisions. As past decisions add up, then, the range of possible interpretations gets increasingly constrained. Similarly, in our own lives, when there is a question about how our values apply to a new case, we might be guided both by thinking about those values and by thinking about what would be coherent with the way we have understood them in the past. Constraining our choices in this way allows us to see our current choices and past ones as expressing a coherent take on the world. Perhaps the reasons we acquire when we commit get their normativity from the value of this kind of continuity.

One way to respond to such a view might be to say that they are illicitly smuggling in

251 Dworkin (1986)

some voluntarist commitments. Maybe the idea that integrity or planning is intrinsically valuable only sounds plausible because we have voluntarist understandings of these ideas in mind. For instance, Chang objects to this kind of externalist strategy on the grounds that, if it were true – if it really were that making a commitment only had normative significance in light of these other values – then no clear-eyed agent could make commitments. Why not? As she develops the argument, the idea is that making a commitment involves taking yourself to be creating a reason. So, if ultimately you're not, making a commitment would involve an act of self-deception.[252]

Chang illustrates this idea by focusing on cases of romantic commitment. Structurally, her argument works like this. She starts by ruling out analyses of this kind of commitment in terms various other psychological kinds, like promises, complex intentions, and the like. Having ruled out these other apparent options, she concludes that commitments like this must involve acts of will. She then draws the further conclusion that romantic commitments should be understood in a voluntarist way.

I'm skeptical about this argument. In particular, I have a hard time seeing what gets us from the intermediate conclusion that romantic commitments involve the will to the main conclusion that the role the will plays in them is metanormative. This seems to be exactly what's at issue between Chang and the Externalist Accommodationist. One could agree that the will is involved somehow, but think that the way it makes a difference is by triggering the antecedent of some normative principle.

So, it seems to me that for Chang's argument to succeed, it would need to be the case that the concepts employed by various versions of Externalist Accommodationist strategies –

252 Chang (2013a)

planning, having integrity, committing – constitutively include a voluntarist understanding of them. That would explain why thinking that you're making a commitment involves thinking that you're creating a reason, rather than just doing something that somehow ends up making a difference to your reasons. But, I don't see why we should assume that. All that seems to be required to commit in a non-self-deceived way is the thought that one's commitments have normative significance. What metaphysically explains that significance seems to be a further matter. Compare: I don't need to know why my promises are normatively significant (maybe it's the disvalue of subverting others' expectations, maybe it's the value of the institution of promising, maybe it's the dignity of the person I'm making the promise to, maybe promising creates reasons for me in a voluntarist way) in order to make promises in a clear-eyed way. All I need to know is that there's this thing I can do (promise) that has these normative effects. Similarly with committing, it seems that there's this thing I can do (engage my will) that has these normative effects. Chang's argument seems only to help us narrow in on what the thing I can do is, not what explains why it has the effects it does.

So, I think, this way of responding to the Externalist Accommodation strategy is not so promising. As far as I can tell, there's nothing inherently voluntarist about the ideas of commitment, integrity, or planning. Still, I think there is another way of responding to this general kind of Externalist Accommodationist strategy. While it's not incoherent, it nonetheless leads to an unsatisfying picture about the relationship between us and our reasons to carry out our commitments – or, at least, a less satisfying picture than the one offered by hybrid voluntarism.

To see why, start by considering the way Brink objects to internalism. Recall from Chapter 3 that one of his arguments for thinking that desires themselves can't be the ground of reasons hinges on the idea that this would make our relationship to our reasons objectionably narcissistic. Consider the reason I have to keep my promises. Intuitively, what grounds this reason is something about *you*, the person I've promised – your rights, legitimate expectations, or authority to call in this favor. In contrast, internalism seems to make the ground of this reason something about *me* – my desire not to let you down, or to be the kind of guy who keeps his promises. This seems to locate the ground of the reason in the wrong place.

I noted in Chapter 3 that one way for an internalist to resist this objection is to distinguish the content of our reasons from the background conditions that explain why they are reasons. But, I also pointed out, doing that robs internalism of some of its theoretical appeal. One interesting feature of this argument is that it appears to be a normative argument for a metanormative conclusion. It would be bad if all of our reasons turned out to be narcissistic – that's why internalism's metaphysics is unappealing even if it gets the content of our reasons right – and that gives us some reason to prefer a theory that doesn't force that conclusion on us. My worry about Brink's version of Externalist Accommodation applies this same argumentative strategy. Roughly, I think a world where some of our reasons are self-created is a normatively better one, and that's some reason for preferring a metanormative view on which that's possible.[253]

253 In general, it's a bad an inference to go from a premise of the form "it would be better if p" to the conclusion "p." It would be better if no one suffered from hunger, but people do. Nonetheless, I think inferences like these are okay in particular circumstances. One thing that makes metaphysical theorizing different from empirical description is that more than one metaphysical theory can be empirically adequate. As a result, when we're choosing between two empirically adequate theories, there's room to appeal to some non-empirical grounds (think: parsimony, fecundity, elegance). One of those grounds, I'm suggesting, is whether the metaphysical

The kinds of cases I have in mind are ones where the meaning of an outcome depends on how that outcome was brought about. For example, in my family, cooking for others is a primary way of expressing your affection for them. So, if I invite a friend over for dinner, it doesn't just matter to me that she has something to eat; it matters to me that I bring it about that she has something to eat. The same outcome, brought about another way – say, if I'd ordered delivery instead – wouldn't have the same significance. Contrast this with other outcomes, where all that matters is that they're achieved. When there's a power outage, it doesn't matter to me who makes the lights come back on, just that they do. Not everyone feels this way about cooking, but the same phenomenon is recognizable in other places. Many of us, for instance, don't just care that we end up dressed in the morning; we care that we dress ourselves.

In the cooking example, the outcome I'm concerned with (there being something to eat) is a state of affairs that doesn't explicitly refer to reasons. But that an agent has some reason or another can also be something one strives to bring about. I might, by practicing my pool game, strive to make myself into the sort of person who has a reason to bet a round of beers on my play.[254] So, I want to suggest, just as it can make sense to care that we be the ones to bring about some state of affairs, it can make sense to care that we be the ones to bring it about that we have some reasons. For example, one might plausibly imagine Cecil as not just caring that he has a reason to give Carlos a kidney, but caring that his commitment be what makes it the case that he has that reason. Carlos might be disappointed, if, say, what explains why he has a reason to give

theory forces bad normative consequences on us. Compare: we might resist an account of what free will (metaphysically) requires on the grounds that if true it would rule out something (normatively) good, like friendship. Thanks to Alex Rajczi for pressing me about this and to Dana Nelkin for prompting me to think about the free will analogy.

254 When I first wrote this example it was purely aspirational. Thanks to Rosalind Chaplin and Emma Duncan for shooting enough pool with me to make it realistic.

Carlos a kidney is that he was randomly drawn from a list of compatible donors. That way of coming to have the reason wouldn't be something he did himself.

On a hybrid voluntarist proposal, it can turn out that our wills are the explanation for why we have reasons like these, while on Brink's proposal, they can do so only with the help of facts about the value of agency (analogously, when I order takeout I only bring it about that there is something to eat with the help of the restaurant). In this way, hybrid voluntarism gives us a more appealing account of our relationship to our reasons. For hybrid voluntarists, making a commitment and thereby acquiring a reason is like getting dressed.

This point parallels one that Frankfurt makes about reasons of love. Though he's not working with a voluntarist conception of the will, he suggests that acting on reasons that are somehow central to our identity gives rise to an "invigorating release and expansion of ourselves."[255] Giving this idea a voluntarist gloss makes sense of what's so invigorating about this kind of experience, and why it would seem like an "expansion" of ourselves. That it's the exercise of a distinctive agential power explains why it's invigorating, much as exercising your power to go for a run can be invigorating if you're otherwise feeling sluggish. That it introduces new reasons explains why it's an expansion; making a voluntarist commitment to a project enlarges the set of things that matter to you and to which you rationally ought to be responsive.

This argument depends on the intuition that it can matter to us that some of our reasons are ones we make ourselves. If so, a metanormative picture according to which that's possible is preferable to one according to which it's not (just as, for Brink, it can matter to us that some of our reasons are not narcissistic, and so a metanormative picture according to which that's

255 Frankfurt (2004), p64-65.

possible is preferable to one according to which it's not). This is particularly clear, I think, in cases where the reasons in question are connected with our personal projects and relationships.[256] But you might not be inclined to give the cases the same gloss, so let me now try to say more in support of this intuition.

One way to elicit this intuition is to consider the value of reflection on what to do. Sometimes we act on reasons more or less on autopilot. Other times, though, it's important to us to reflect on them. For instance, while it would be fine to go through life relying exclusively on friends' movie reviews, something would be missing if we made decisions about our careers or romantic partnerships this way. In cases where we're deciding on big projects around which we will organize other parts of our lives, we want not just to respond to the reasons that there are, but to appreciate what those reasons are and how they fit with the rest of our sense of ourselves. So, there is an asymmetry among cases of action for reasons; sometimes we want just to respond to reasons and sometimes we want to stand in some more robust relation to them.[257]

Intriguingly, this asymmetry tracks the one we encountered in Chang's argument about continued deliberation. Sometimes we're content to pick arbitrarily, while sometimes continuing to deliberate seems appropriate. Picking arbitrarily among sufficiently good options seems appropriate in cases where all that matters to us is responding to reasons, while continued deliberation seems appropriate in cases where we want this more robust relation.

256 There is some debate about whether the reasons we have in connection with our personal projects and relationships essentially involve agent-relative reasons, or whether they can be explained in terms of agent-neutral ones. I think it's the former, but that's not the point I'm trying to make here; after all, externalists can recognize agent-relative reasons just as well as voluntarists can. I'm claiming that sometimes it matters to us how we acquired the agent-relative reasons that we have in these cases. That's the observation that I take to support hybrid voluntarism. See Perry (1976) and Whiting (1986) for discussion..

257 In Davia (2018), I argue that this kind of investment in the reasons for which we act is a constitutive part of being an agent in the first place.

On non-voluntarist views, it's a a little bit mysterious what this more robust relation amounts to. Unless reflection can create new reasons, it's not clear that having reflected on some reason makes it any more a part of who we are, or makes our acting on it any more of an achievement. After all, it seems that something can be part of who I am without my having reflectively endorsed it. In contrast, on hybrid voluntarist views, there's a natural candidate for what we're after when we want this more robust relation to our actions, and why reflection seems like the way to get it: we sometimes want to act for reasons that we have played an active role in bringing it about that we have, and reflection is the way we play that active role in creating reasons. So, if we explain the will's normative significance in terms of its metanormative significance, we get the resources to explain why reflective choice, rather than just choice, is sometimes important to us.

Another way to get at this intuition is to look to work it seems to be doing in other areas of philosophy. Alex Rajczi, for example, appeals to an idea much like this one in an argument against consequentialism.[258] Rajczi observes that, in a world where consequentialism is true, what values are to be maximized are determined independently of agents' decisions about what to do. So, agents will only have genuine, morally permissible options in cases where the consequentialist calculus produces ties. In contrast, in a world where consequentialism is not true, there is room for the possibility of agents making some option valuable by choosing it. This, Rajczi thinks, is itself valuable, and so there is a kind of value that's possible in the non-consequentialist world but not in the consequentialist world.[259] Here, Rajczi is just concerned

258 See Rajczi (2011).

259 Why can't the consequentialist just include this value – the value of making something valuable by choosing it – in the set of values to be maximized? Rajczi answers: because the valuable the thing here is in what brings about an outcome, rather than whether or not the outcome is brought about.

with a dispute about first-order morality (is it consequentialist in form or not?) but a metanormative intuition like the one I'm suggesting here seems to do work for him. This suggests, I think, that whether or not one finds this idea plausible isn't just a matter of having prior externalist or voluntarist sympathies.

The idea that direct control over one's reasons is a thing it makes sense to want also shows up in the free will literature. Several philosophers have thought that there's a kind of freedom worth wanting that occurs within "torn decisions," i.e. cases where we have sufficient reason to do either of two things and what we'll ultimately do depends on which options we put our will behind.[260] In these situations, these philosophers suggest, we exhibit a kind of freedom distinct from the kind we have when we merely respond to the reasons that there are. These philosophers typically take themselves to be concerned with the kind of free will required for moral responsibility, and I am skeptical that this kind of freedom is required for that, but I nonetheless think these philosophers are on to something. Even if control over one's reasons is not worth wanting because of its contribution to moral responsibility, it could nonetheless be worth wanting as a distinctive expression of our agential powers. If there are any voluntarist reasons, then this kind of freedom is one we really have.

One can find a similar idea in the literature on promising. David Owens, for instance, suggests that in many cases of promising, we have an interest not just in what happens, but in who has authority over what happens.[261] Imagine, for instance, that I ask you to pick me up at the airport. You respond by saying that you're very likely to pick me up, but that you won't promise

260 See, for example, Balaguer (2010) and Kane (2012) for different ways of developing this idea. See also Nozick (1983), Chapter 4.
261 See Owens (2012), Part Two.

to do so. This should worry me. Moreover, it should worry me even if I know that your reluctance to promise is not explained by a suspicion that you won't really be able to pick me up. Cases like this suggest that what we want in promise-situations is not just to get what we're promised (otherwise, knowing how likely you are to pick me up would do the trick), but also to have authority over whether we get what we're promised. If you promise me, I have the authority to demand being picked up in a way that I don't if you just reliably expect to be available. Something similar, I think, is true of the ability to create voluntarist reasons. Cecil might want both to have a reason to give Carlos his kidney and to be in a position of authority over whether or not he has this reason.

So far I've tried three ways of making it plausible that it sometimes matters to us that our reasons can be self-created: this idea seems to be lurking in our thinking about the value of reflection, arguments against consequentialism, and arguments about the kind of free will worth wanting. This intuition also gets support from thinking about different ways we can experience our hobbies or personal projects.

When we're thinking about what we're doing, we can take either a *telic* or *atelic* perspective.[262] For example, when thinking about writing a dissertation, I might focus on a future goal that this activity contributes to (say, that it's a necessary step toward getting an academic job) or I might focus on features of the activity as I'm experiencing it in the moment (say, that doing philosophy is fun). Some activities are such that one perspective is easier to inhabit than the other. For example, lifting weights is typically experienced in a telic way; one lifts in order to

262 My use of this contrast was inspired by Setiya (2017).

get stronger.[263] Playing catch, by contrast, is typically experienced in an atelic way; one plays just to play, rather than in the service of some further goal.[264] Many activities, though can be experienced either way, depending on what one chooses to focus on. For instance, when I brew beer, I typically approach the activity in an atelic way – I just find it fun to go through the procedures – but I sometimes consider taking steps to adopt a more telic perspective. I consider, for instance, taking notes on past recipes, controlling more carefully for variables like the ambient temperature, investing in more precise tools, etc. These things would help me direct current brewing activities toward future brewing goals.

In cases where an activity might be experienced in either of these ways, our reasons for participating in the activity can influence which experience we end up having. Once we have a reason in mind, we can ask whether we're engaged in the activity in a way that is guided by that reason. For instance, if what justifies my brewing beer is that later we'll need some beer for a party, this sets constraints on what counts as a good brewing performance. I should make brewing decisions (say, whether to add the hops earlier or later in the boil) in light of the the goal toward which the brewing is directed (say, how bitter the guests like their beer). As a result, brewing in order to provide beer for a party will tend to feel different from brewing just to do it; the former involves constraints that the latter does not.

This point – that conceiving of an activity as done in the service of a further goal can change how it feels to do it – helps bring out the intuition I've been defending here – that it can matter to us whether some of our reasons are self-created. Plausibly, a good life involves some

263 The "typically" qualifier is important. I suspect that Michael McKenna and Justin D'Ambrosio lift weights just to express their agency, and that's just drawing from philosophers I know.

264 Other paradigm examples include things like going for a walk or chatting with friends.

activities that are experienced in an atelic way. So, a good life involves some activities that are done for reasons that don't set constraints on what counts as a good performance of the activity (think of playing catch: beyond merely being able to throw and catch it's not clear what would even be involved in being better or worse at it[265]). But, since the reasons for participating in an activity set constraints on what counts as a good performance, it can be a bit puzzling how we can both be justified in a pursuit and not impose telic constraints on it. Voluntarist reasons fit the bill nicely here. If my reason for brewing is explained by my commitment to it, any way of brewing will do the job. In contrast, if my reason for brewing is explained by my commitment to it and the fact that following through on my commitments respects the value of agency, there's room to wonder whether my performance is doing better or worse at respecting the value of agency, just as there's room to wonder whether my performance will be better or worse at providing drinks for the party. Maybe it would be better, from the perspective of exercising my agency, to opt for brewing projects that require more sophisticated planning, more practice with new techniques, or more complex coordination with others. Brewing just to do it does not introduce these considerations.

This is not to say that experiencing activities in a telic way isn't valuable too.[266] The point here is just that it's valuable to have some parts of one's life that one experiences in an atelic way,

265 Maybe limiting dropped catches is a better performance of catch? Or throwing harder or more accurately? I doubt it, at least once you're above the threshold where you can throw and catch well enough to play. Someone who tracked these things or tried to maximize them would be seriously misunderstanding the activity.

266 In this way, I disagree with Aldo Leopold's view that “becoming serious is a grievous fault in hobbyists.” See Leopold (1972), p4. Nonetheless, my thinking about what's distinctive about the value of activities we experience in an atelic way was inspired by Leopold. I got started thinking about this after posting a quote from Leopold on Facebook and getting objections from David Brink, Sam Rickless, and Michael McKenna. So, thanks to them for pushing me to figure out what, despite their good objections to his strong claim, I found attractive in Leopold's idea.

and it's easiest to see how we can do this if our reasons for engaging in those activities are up to us. Otherwise, whatever our reasons for engaging in the activity are threaten to impose a telic structure on it. So, if it matters to me that I have some atelic experiences, it should matter to me whether some of my reasons are self-created, because engaging in activities for those reasons facilitates my experiencing those activities in an atelic way.

In this section, I have tried to argue that hybrid voluntarism provides a more satisfying explanation of the normative significance of the will than do Externalist Accommodationist views. The core of my suggestion was that sometimes it makes sense to care about playing an active role in bringing it about that we have the reasons we do. On hybrid voluntarists views, doing so is possible, while on externalist views, we can only bring it about that we have the reasons we do with the help of the externalists facts. Why accept that it makes sense to want this kind of active role in bringing it about that we have particular reasons? The intuition that it does turns out to play a role in the value of reflection, first-order ethical arguments about consequentialism, the kind of free will worth wanting, and the value of pursuits we experience in an atelic way.[267]

## 3. Taking Stock

If what I've said so far is right, then hybrid voluntarism has advantages both over views that deny the normative significance of the will and over Externalist Accommodationists who

267 Note that my goal in pointing to these other arguments is to bring out the plausibility of the intuition that it can matter to us whether some reasons are self-created. So, all I need is that the intuition plays a role in motivating those arguments. I don't need the arguments to ultimately work out. For example, maybe consequentialism is true after all, or maybe torn decisions don't really tell us anything about free will – I don't claim to settle any of that here. My point is just that, even if those arguments fail, they are initially appealing, and their initial appeal sheds light on our desiderata for metaphysical theories about reasons.

accept it. As far as I can tell, these arguments are available to defenders of any version of hybrid voluntarism, so long as they don't fall prey to the extensional problems Korsgaard's purebred voluntarism does. For instance, I think Chang can take them on board, as could versions that relax her constraints in the ways I explored in Chapter 5.

Stepping back a bit, let me note one methodological argument for putting some weight on these benefits of hybrid voluntarism when comparing it to other theories. In theory choice, it's a good thing if a theory has fecundity, i.e. if it suggests new ways of looking at old problems, and if it has the resources to address questions that it wasn't initially designed to answer. The arguments I've made in this chapter suggest that hybrid voluntarism has that virtue. As we've seen, the main proponents of voluntarism have tended to be focused on issues about identity and the authority of reasons. These are important questions, and I think hybrid voluntarism does make a contribution to understanding our relationship to our reasons. But, if what I've said here is right, the voluntarist idea also helps to explain other phenomena. These aren't the puzzles that voluntarism was introduced in order to solve, so that voluntarism offers us insight into them gives us reason to think it's on the right track, and to be optimistic that the hybrid voluntarist idea might have other helpful applications.[268]

The arguments I've made, then, provide us with reasons to accept a view about the metaphysics of normativity according to which some normative facts are brute, but others are

268 One idea for a further application is in legal philosophy. One might think of the authority of law in terms of laws having been chosen by the population they govern (e.g. in hypothetical contract), or one might think of the authority of law in terms of the valuable ends the law serves. We might see the former of these approaches as a kind of voluntarism at the collective level. But just as in the individual case, it might be that this way of thinking about the options is too narrow: perhaps there a different kinds of legal authority, or different laws get their authority in different ways. Obviously a lot more would need to be said about how a hybrid voluntarist view about practical reasons might extend to a hybrid account of legal reasons, but it's a project I'd like to explore in the future. Thanks to David Brink for pointing out the possible analogy here.

explained in terms of agents' wills. We need to recognize the brute facts in order to avoid the extensional problems of purebred voluntarism or purebred internalism, and recognizing the voluntarist facts helps explain the phenomena I've been discussing in this chapter. In particular, the non-externalist portion of a hybrid view should be voluntarist, because only voluntarism explains how it could be that we play an active role in determining what reasons we have. This contrasts with a hybrid between externalism and internalism, on which it would be true that an agent's mental states are part of the explanation for the reasons she has, but it would not be true that her reasons are up to her. At least in the moment, we just have the desires we have.

If one accepts the arguments for recognizing voluntarist reasons as part of a hybrid view that I've been making here, one might wonder: why not go further and think that there are *three* sources of normativity? Perhaps some reasons are brute, some are explained by our wills, and some by our desires. I haven't offered any arguments against that view here – it certainly wouldn't have the extensional problems of purebred internalism nor, thanks to its voluntarist component, would it have trouble explaining how we can have an active role in determining our reasons. So, for all I've said here,  that view might be right. Since my primary concern here is to argue that we ought to recognize some voluntarist reasons, I'm happy to concede this possibility.

Nonetheless, such a view seems unmotivated. The considerations I've offered on behalf of voluntarism in this chapter seem to count both in favor of purebred voluntarism and in favor of hybrids with a voluntarist component, while the considerations typically offered in favor of internalism (e.g. those resting on skepticism about externalism) seem to count only in favor of purebred versions. Once we've admitted that there are some brute normative facts, and once

we've allowed that for others some kind of metanormative explanation is possible, it's hard to see what further work there is for an internalist component of a hybrid view to do. So, while I haven't ruled out this possibility, I think there is a *pro tanto* case for thinking that of the hybrid views, a hybrid between externalism and voluntarism only is the most promising.

In any case, the main work of this chapter was to get some different versions of hybrid voluntarism on the table and to explain why one might be attracted to a view in that family. Let me close by saying a bit about which of these versions I find most plausible.

My preferred view is one that relaxes both of Chang's constraints, accepting *Indeterminacy or Proximity* as an account of when it's possible for agents to create reasons and *No Dramatic Valence Changes* as an account of what normative consequences those reasons can have. For short, let's call this the *Permissive View*.[269] One benefit of the Permissive View has to do with the argument from normative motivation we considered in Chapter 5. In the course of making that argument, I noted that hybrid voluntarism only plausibly helps explain normative motivation if the will has normative significance in a both (a) a wide range of cases and (b) cases where we're deciding things that are important to us. Since relaxing both constraints makes the will normatively significant in a wider range of cases and puts weaker restrictions on the effects it can have in those cases, (a) and (b) are more clearly satisfied on that view.

This same feature of the Permissive View may also make it better-positioned than Chang's to respond to the worry about practical guidance that Raz levels at purebred voluntarism.

269 As I noted in Chapter 5, even if go for the Permissive View, there are some other choice points for hybrid voluntarists, e.g. whether some domains of reasons (e.g. moral or prudential) are subject to special constraints. Another choice point has to do with the role voluntarist reasons play in deliberation. Perhaps they are just some among many reasons in the hopper, or perhaps they sometimes have a preemptory force, ruling out the consideration of other reasons that otherwise would have been relevant. I'm officially agnostic about those questions here, but hope to explore them in future work.

Recall that Raz worried that if there are reasons whose existence is up to us, we will be immune to rational criticism when we fail to respond to those reasons because, were criticism advanced, we could always just take back the commitment that grounds the reason we're failing to respond to. On behalf of Korsgaard, I noted that just because one *can* change one's commitments doesn't mean one *actually does* so, and in the cases where one doesn't one is still on the hook for rational criticism.

The Permissive View can improve on this reply. As I've put it so far, this move depends on the (plausible, I think) assertion that sometimes we fail to live up to our commitments without thereby abandoning them. But Raz might think that this is exactly what voluntarists need to explain – why not just change your commitments any time doing so makes your actions come out more rational? In light of the arguments I made above, hybrid voluntarists can answer that creating will-based reasons is part of a full expression of our agential powers. So, by giving up the commitments that give rise to those reasons at the drop of a hat, we would fail to be fully exercising our rational agency. And, given that the Permissive View allows us to do this in a wider range of cases than Chang's does, the Permissive View makes it more plausible that this would be a serious loss.

Another source of appeal for the Permissive View concerns thresholds and boundaries. If we accept either *No Valence Changes* or *Only Indeterminacy*, our view will have to recognize cases where a small non-normative difference makes a big normative difference. *Only Indeterminacy* draws a sharp line between cases of parity, ties, and indeterminate inequality on the one hand, and cases of determinate inequality on the other. Cases near the boundaries will be

ones where a small non-normative difference can make a big normative difference (specifically, whether or not we can create reasons). Similarly, *No Valence Changes* draws a sharp line between cases of parity and ties on the one hand, and cases of indeterminate inequality on the other. Cases near the boundary between these will be ones where a small non-normative change to the case will make a big normative difference (specifically, whether or not a reason we create can change what we ought to do).

On their own, these observations about small non-normative differences having large normative consequences are not necessarily objectionable; sometimes small changes have big effects. For instance, if a trolley is going to hit me, you have a *pro tanto* reason to turn it away. But if we make a small change to my location, such that the trolley is only going to zoom past me to the left, you have no such *pro tanto* reason. Similarly for cases where a small change unlocks an emergent property: for a basketball team, getting a slightly better three-point shooter might make make the post-players much more effective. Nonetheless, as a general matter of philosophical methodology, I think it's best to introduce thresholds only as a last resort. Sharp boundaries call out for an explanation of why the boundary should be sharp (trolleys stay on their tracks, defenders have to stay closer to their mark). So, given that the Permissive View seems to respect the extension of reasons without introducing this explanatory demand, I think that counts in its favor.

The Permissive View also gains some plausibility by making room for normatively significant conscientious objection. It's a consequence of both Chang's view and the Permissive View that sometimes we can create reasons even when those reasons don't ultimately determine

what we ought to do, i.e. when they're outweighed by our antecedent reasons. But the Permissive View allows this in a wider range of cases than Chang's does, as it says that we can create (weak) reasons even when our antecedent reasons deliver a wholly determinate verdict. And this, I think, is the result we should want. To adapt an example from Williams,[270] suppose that you work at an engineering firm that is sometimes hired to work on weapons systems. When you first took the job, you had no objection to this, but after working there for a while you change your will; you no longer see yourself as the sort of person who can contribute to a war effort. This might not change what you ultimately ought to do, perhaps your family is depending on your income. But whether or not you can change what you ultimately ought to do, I think it matters whether changing your will this way can give you a (ultimately outweighed) reason to quit. Knowing that your objections to your company's clients matter, even if they don't matter as much as your obligations to your family, might be part of what helps you go on despite them.[271]

Finally, if one starts out finding Chang's view plausible, I think there is some pressure to move toward the Permissive View. Consider the case of determinate ties. If you think that the we can create reasons in these cases, then it's hard to see why we shouldn't be able to do this in cases that are close to ties. Otherwise, we seem forced to think that self-created reasons are very weak ones. But that's just to say you should be tempted to reject *Only Indeterminacy*. Similarly for the strength of reasons, if you think self-created reasons can change what you ought to do in a tie but not in near-ties, it looks like you have to think that self-created reasons are very weak. So, once

270 See Williams (1973).

271 This parallels the point I made in Chapter 5 about why it might matter to us that we be able to resolve a justificatory tie at the time of action, rather than just to introduce new reasons downstream. Having a particular reason can make a difference to how we ought to feel about a decision even if it doesn't make a difference to what we ought to do.

you deny *Only Indeterminacy*, you should be tempted to deny *No Valence Changes* as well. Once you do, you end up at the Permissive View.[272]

These arguments may not be decisive; maybe there are other explanatory benefits of Chang's particular version of hybrid voluntarism, or maybe there are other ways of fiddling with the constraints that produce different versions of hybrid voluntarism. Nonetheless, I think that once we see that Chang's constraints can be relaxed without threat to extensional adequacy, more restrictive versions of hybrid voluntarism start to look unmotivated.

272 Thanks to Dana Nelkin for the idea that if you think the will is normatively significant in ties, there is pressure to think it's normatively significant in approximate ties as well.

# Chapter 7

## Playing Defense and Tallying Plausibility Points

Over the last two chapters, I've tried to identify what's appealing about a hybrid voluntarist account of reasons, first by defending some arguments for it already in the literature and then by offering some new ones of my own. But identifying the appeal of a view falls short of establishing that we should accept it; even if hybrid voluntarism has unique explanatory power, that power might come with other theoretical costs. So, in this chapter, I'll play defense. First, I'll look back at some reasons for skepticism about purebred voluntarism that we've encountered already, asking to what extent they pose a problem for hybrid voluntarism. Those objections target the *voluntarist* part of hybrid voluntarism. Second, I'll consider some objections having to do with theoretical unity that target the *hybrid* part.[273] Finally, I'll step back and ask how this tallying of theoretical costs and benefits compares with the views I engaged with in earlier chapters.

### 1. Worries about Voluntarism

In Chapter 2, my primary complaint about Korsgaard's voluntarism was that it doesn't get the extension of reasons right. However, I also considered some non-extensional objections, i.e. worries about whether the voluntarist idea is plausible even in the cases it gets right. If those objections succeed, they would seem to pose just a problem for hybrid voluntarism just as much as they do purebred voluntarism. So, we ought to revisit those objections and see to what extent they raise trouble for hybrid voluntarists.

273 Behrends (2015) also surveys and responds to some objections that target hybrid voluntarism specifically rather than voluntarism in general, though I think the ones I engage with here are more pressing.

One set of objections to Korsgaard had to do with the the contingent and revisable status of reasons we create by will. If our reasons depend on our wills, and we can change our wills, then we can change our reasons. As a result, one might worry about whether those reasons can provide us practical guidance, be authoritative for us, or ground rational criticism when we fail to act on them. The same kind of objection might be raised against hybrid voluntarism. Even if only some of our reasons depend on our wills, it'll still be true that *those* reasons are contingent and revisable.

Fortunately, I think the reply I offered on behalf of Korsgaard to this set of worries works for hybrid voluntarists too. *Any* view is going to recognize some contingency and revisability; what we do can change what's the case, and changes to what's the case can change what we have reason to do. So, it can't be that these features, all on their own, make voluntarism objectionable. For there to be an objection that voluntarism faces but other views do not, we would need to identify something particularly objectionable about the kind of contingency or revisability that is generated when we cut out the middle step and say that what we decide can change our reasons directly, rather than by way of downstream effects on what's the case. But, I argued in Chapter 2, it's hard to see how to do this without begging the question against voluntarism. Moreover, as I argued in Chapter 6, hybrid voluntarism offers resources for a more forceful response to the worries about practical guidance and rational criticizability.

Nonetheless, I didn't argue that *all* of the non-extensional objections to Korsgaard are question-begging. One issue that I left unfinished in Chapter 2 is whether the phenomenology of deliberation tells against voluntarism. When we decide what to do, we're typically thinking about

the world, rather than about our commitments or practical identities. So, voluntarism might seem to suggest an implausible account of how deliberation works. In Chapter 2, I suggested a two-pronged reply purebred voluntarists can take toward this worry. First, they can note that what metaphysically explains why some consideration is a reason need not be part of the content of that reason. So, it could be that what we should be thinking about when deliberating are just particular facts about the world, even if the metaphysical story about why we should be thinking about those facts appeals to our psychology. Second, purebred voluntarists can note that deliberation is a mixed bag sometimes it does seem appropriate to look inward. So, this objection might oversimplify the data to be explained. With the machinery of Chapters 5 and 6, we can put this second point in a less defensive way. I've argued that deliberation is sometimes inward-looking, e.g. when reasons run out or when we want not just to act on reasons but to stand in some more robust relationship to them, and hybrid voluntarism helps explain when and why this kind of inward-looking deliberation is appropriate. So, all things considered, I think attention to the phenomenology of deliberation supports rather than undermines hybrid voluntarism.

Another worry was about whether voluntarism makes our reasons objectionably unstable. If it turned out that always having the option to change our reasons prevented us from treating any of them as settled when making decisions, deliberation could never get started and we really would have a problem for voluntarism on our hands. In Chapter 2, I noted that this objection bears an interesting relationship to the extensional objections to Korsgaard. If Korsgaard were right that agency constitutively involved some constraints on our willing with substantive consequences, there would be some reasons that aren't always up for revision. That's only so

comforting, though, since I think Korsgaard is wrong about those constitutive requirements. Nonetheless, this observation does open the door to a more satisfying response to this worry on behalf of hybrid voluntarists. Hybrid voluntarists don't have to rely on constitutive claims about agency in order to recognize reasons that we have categorically; the externalist part of the hybrid can supply us with those. So, even if it's true that all voluntarist reasons are in principle revisable, and even if it's true that this revisability necessarily interferes with deliberation, hybrid voluntarists aren't forced to conclude that deliberation doesn't get off the ground. And so long as there are *some* fixed points for deliberation to get started, that other commitments are revisable does not seem so implausible.

The same applies to the classic Euthyphro worry that voluntarism makes our reasons objectionably arbitrary. I noted in Chapter 2 that this objection is at it's most forceful when construed as the worry that *all* of our reasons are arbitrary, the way they would be on a purebred voluntarist view, or at least the position a purebred voluntarist falls into if Korsgaard's arguments for moral rationalism don't hold up. Thankfully, hybrid voluntarists needn't pin their hopes of responding to this objection on those constitutivist arguments. Hybrid voluntarists can instead point to the externalist part of the hybrid, and note that pockets of arbitrariness within a broader system of norms were never objectionable (think again of driving: it's arbitrary that we drive on the right rather than the left, but not arbitrary that we drive in some way or another designed to reduce traffic accidents).

So, hybrid voluntarism seems unthreatened by the objections we encountered when considering Korsgaard's purebred voluntarism. But, as we saw in Chapter 3, there are non-

extensional worries about internalism too. In particular, one might worry that the narcissism problem – the idea that internalism makes the ground of our reasons objectionably self-regarding – applies to to voluntarist views just as much as it does internalist ones.

Leveled against a purebred version of voluntarism, I think this is a good objection. Some reasons seem not to depend on us, and our theory of the metaphysics of reasons should accommodate that intuition. And it's possible that this objection has purchase against some versions of hybrid voluntarism as well. If, for example, whether I had reason to care for others just depended on facts about me, that might be an objectionably narcissistic theory about those reasons. But, this doesn't seem to be the case with the versions of hybrid voluntarism we considered in Chapters 5 and 6. There, the opposite is true: decisions about life-organizing commitments like careers, romantic partnerships, long-term projects, and so on seem like exactly the sorts of things that should be grounded in us and our choices. So, while this objection might threaten the view that all of our reasons have a voluntarist source, or that a subset of reasons like our moral reasons have a voluntarist source, it doesn't threaten the kinds of views I've been exploring here.

All in all, then, hybrid voluntarism seems to avoid the objections leveled at purebred views. It might, however, give rise to new problems of its own. One worry concerns supervenience. Plausibly, the normative supervenes on the non-normative, in that there can't be a change in the former without a change in the latter. Maybe hybrid voluntarism runs afoul of this requirement? One way in which it comes close is that it allows changes in our wills to have normative effects, even when other non-normative facts are held constant. So, for example, if my

doppelganger commits to learning Spanish, he'll have a reason to do so that I don't, even though all the other non-normative facts about us are the same (like, that we live in cities with large Spanish-speaking populations, that speaking Spanish would make some kinds of travel easier, etc). Of course, in this case there is still some interesting non-normative difference between me and my doppelganger: the difference in our wills. So, there's no strict violation of supervenience here.

One might think that this move – including the facts about our will in the supervenience base – isn't really available to hybrid voluntarists. Perhaps it collapses into to the Externalist Accommodationist proposal we considered in Chapter 6, where there are normative facts of the form "if you will such-and-such, you have reason to do thus-and-so." Reasons we create by will, this thought goes, are not the kind of reasons we get just by being in particular circumstances (like our reasons to be charitable if we're wealthy, or attend talks if we're members of the department). So, perhaps it's illegitimate to include them in the set of circumstances that the normative is supposed to supervene on. And if they're excluded, it looks like voluntarists might have to deny supervenience after all.

This worry turns on the idea that, if we include facts about our wills in the supervenience base, then sentences like "if you commit to learning Spanish, then you have a reason to do so" will come out true, and that this will collapse hybrid voluntarism into a kind of externalism.[274] But it's worth remembering from Chapter 6 and our discussion of the Externalist Accommodationist that sentences like this admit of two readings. We might read them *normatively* in which case they express the content of some normative fact (if you commit to

274 Thanks to Ruth Chang for pressing me on this.

learning Spanish, then you have a reason to do so), or we might read them *metanormatively* in which case they express an explanatory relation between some non-normative fact (you have committed to learning Spanish) and a normative fact (you have a reason to learn Spanish). Given that there are these two readings, hybrid voluntarists and externalists can agree that there are true conditionals like this, but disagree about what makes them true. So, agreeing that facts about our wills are some of the facts that the normative supervenes on needn't amount to agreeing that this supervenience is explained by an externalist principle. If that's right, hybrid voluntarists can distinguish agent-relative reasons of circumstance and self-created reasons without denying the supervenience of the normative on the non-normative.

As far as I can tell, then, hybrid voluntarism doesn't lose any plausibility for involving a commitment to voluntarism. It might, however, lose some plausibility for being a hybrid. So, let's turn our attention to that possibility.

## 2. Worries about Hybrid Views

One risk for any hybrid theory is that it combines two views in an objectionably ad hoc way. Maybe the versions of hybrid voluntarism I've suggested avoid extensional objections, but only at the cost of making for a theoretically clunky view.

In assessing this risk, it's worth remembering that not just any disjunctive feature of a theory makes it *objectionably* ad hoc. For instance, in baseball a play can be scored as an error either because a fielder failed to field a ball that one could reasonably expect would be fielded, or because a fielder makes a wild throw that could not reasonably be expected to be caught at its intended destination. But no one would complain that this disjunctive feature makes the

definition of an error ad hoc, since both disjuncts really are plausibly ways of making an error. So, hybrid voluntarism doesn't lose plausibility just by being a hybrid; it loses plausibility only if it's a hybrid where the component parts do not fit together. So, to address this worry we need to look at paradigm cases of considerations counting as reasons for some agent, and see whether those cases have this disjunctive feature. If so, hybrid voluntarism won't be objectionably ad hoc, it'll just be disjunctive in the way that the scoring of an error is disjunctive.

Recall Schroeder's case of Ronnie and Bradley from Chapter 3. Schroeder sets things up as follows.[275] There will be a party tonight, and everyone is invited. Ronnie likes dancing, while Bradley hates it. The fact "there will be dancing at the party" seems like a reason for Ronnie to attend, and a reason for Bradley not to. Schroeder goes on to argue that it's Ronnie's desire to dance and Bradley's desire not to that best explains this difference in Ronnie's and Bradley's reasons. I resisted that conclusion in Chapter 3, but I nonetheless think that Schroeder's contrastive method is useful here. Revisiting some variants of this case can help show why hybrid voluntarism's combination of views is not objectionably ad hoc.

First, consider different ways in which Ronnie might want to go dancing. It could be that this is pretty much an unbidden desire, one that Ronnie feels the tug of but doesn't identify with (he might regard it as a distraction from his studies). It also could be that this desire plays a role in a self-description Ronnie values (maybe he thinks of dancing as expressing his inner romantic). It seems to me much clearer that Ronnie has a reason to go dancing in the latter case. Now, consider different desires that might give Ronnie a reason to go to the party. It could be that he likes dancing, or it could be that he likes having the opportunity to make a mess of the

275 Schroeder (2007b) p1-2

host's house. I'm much more confident that the former provides a reason than the latter. Taken together, these variations suggest that our judgments about reasons are disjunctive in just the way that hybrid voluntarism suggests. We see some reasons (say, not to make a mess of the host's house) as not depending on us, and we see the reasons that do depend on our psychology as depending on our relationship to those psychological states. So, while hybrid voluntarism might offer a surprising account of the metaphysics of reasons, it nonetheless tracks commonsense distinctions we make about the extension of reasons. Indeed, as we saw in Chapter 6, these are judgments about what reasons we have that externalists can agree with, even if they disagree about whether the will has any special role in the metaphysics that explain them. It seems, then, that reasons are like baseball errors, the sort of thing for which a disjunctive definition is potentially appropriate.

This is not to say that these variants of the Ronnie and Bradley case directly support hybrid voluntarism; Schroeder might try to explain them in terms of differences in Ronnie's desires, or an externalist might explain them in terms of the value of what's desired. The point here is just that, whether or not these different views give the will metanormative significance, they should converge on the idea that it sometimes has normative significance. But then, if this normative claim is true, the hybrid voluntarist account shouldn't strike us as objectionably ad hoc for making a normatively-relevant distinction.

There is a related objection in this ballpark, though. One might agree that it's not unprincipled to recognize two metaphysical grounds of reasons, but think that doing so comes at the cost of giving up other important desiderata for a theory about reasons. For instance,

Schroeder suggests that if our theory of reasons is disjunctive, we'll be saying that there's really nothing that makes reasons what they are. On a view like hybrid voluntarism, Schroeder says, reasons would be like pieces of jade: lumped together in the extension of some term but without anything shared that explains why they all belong in there.[276] So, while hybrid voluntarism might not be ad hoc, it might still lose some plausibility for not being a fully unified theory. And if one values theoretical unity highly enough, that cost might outweigh the costs of competing theories.

As a start toward a reply, it's worth noting that Schroeder's analogy undersells what can be said about what pieces of jade have in common. While it's true that they have no common physical structure, they all do some share some outward physical properties (e.g. their distinctive color), historical properties (e.g. being used by us for jewelry), and functional ones (e.g. that they can be sold for such-and-such a price, or displayed to communicate such-and-such style). Similarly, even if one goes for a hybrid view about what reasons are, there are plenty of things to say about how all reasons are related: they're the kinds of things we think about when deliberating, they count in favor of and against options, they're the sort of things that tend to motivate us when we're rational, etc. In both cases, there's disunity at the level of individual instantiation, but unity at the level of functional description.

This kind of observation isn't comforting to someone who wants a unified theory of jade, but I think it should be comforting in the case of reasons. The difference is that jade is a physical thing, so you'd expect a physical explanation of what makes it what it is. So, the non-physical things pieces of jade have in common don't help us come up with a satisfying metaphysics of jade. But the concept of a reason is a functional concept rather than a physical one. The

276 Schroeder (2007b), p60

commonplace truths about reasons (that they count in favor of and against options, that deliberating well involves attending to them, that they motivate rational agents) are all about the functional role reasons play rather than their individual features. If so, these kinds of properties held in common are just the right kind to tell us what reasons are. So, it seems to me, unity at the level of a functional description is all we should have expected for reasons. It might have been interesting if a theory could provide unity at the level of individual instantiation too, but failure to do so certainly doesn't mean failing to offer any unifying analysis at all.

## 3. Tallying Plausibility Points

My work in this chapter has been primarily defensive, so it might be helpful to step back and see where these moves leave us. Macroscopicly, my arguments in this dissertation have gone like this. I argued in Chapters 2 and 3 that purebred versions of internalism and voluntarism face problems accounting for the extension of reasons. There are various strategies defenders of those views have tried to solve those problems, but I don't think they work out. Moreover, I argued in Chapter 4, the standard objections to purebred externalism that motivate these views are not decisive. This means that the best view about the metaphysics of reasons is going to recognize some normative facts that are not explained in terms of the desires or wills of the agents for whom they are reasons. One kind of view that meets that desideratum is purebred externalism; another is hybrid voluntarism. And, I argued in Chapters 5 and 6, while both of these views can accommodate our intuitions about the extension of reasons, hybrid voluntarism does so in a more theoretically appealing way. So, unless hybrid voluntarism introduces new theoretical costs, we should accept it. In this chapter, I've argued that it doesn't introduce those costs. So, if all that is

right, we should accept hybrid voluntarism.

This overall argument structure amounts to a tallying of plausibility points: internalism and voluntarism lose some – when compared to externalism and hybrid voluntarism – for their extensional problems, and hybrid voluntarism gains some – when compared to externalism – for its explanatory power. One limitation of this kind of argument is that there may be more than one reasonable way of weighing different theoretical costs and benefits. While philosophical arguments might make clear that something is or isn't a cost or benefit (say by identifying some explanatory power or showing how a theory can avoid a bad implication), there are further questions about how much weight to put on those costs and benefits when comparing them against one another. I am skeptical that these metaphilosophical questions have uniquely good answers, so let me note two places where someone might reasonably do this tallying differently than I have.

First, while I argued above that hybrid voluntarism doesn't introduce any theoretical costs for being a hybrid view, I did leave open the possibility that a theory could do better by being more unified. Specifically, hybrid voluntarism allows us to preserve unity at the level of a functional description of what reasons are, but not at the level of individual instantiation of reasons. Some reasons will be considerations that are normative in virtue of the externalist facts, and others will be considerations that are normative in virtue of our wills. In contrast, on a purebred externalist view, there's only one source of normativity. So, if one puts a lot of weight on theoretical unity, one might grant everything I've said here but still prefer an Externalist Accommodationist view like Brink's or Dorsey's over a version of hybrid voluntarism like mine

or Chang's.

I don't have a knockdown argument against going this way – I'm open to the idea this is one of the places where there's reasonable metaphilosophical disagreement and argument just runs out – but here's why I'm inclined not to. I think of the virtue of theoretical unity as analogous to the virtue of parsimony. Valuing parsimony doesn't lead us to go for the least ontologically-committed theory come what may, only the least ontologically-committed one that's adequate to explain the phenomena. Similarly, I think, we shouldn't leave explanatory power on the table in the name of theoretical unity. Otherwise, what we get is not so much unity as oversimplification. So, if my arguments for the explanatory power of hybrid voluntarism from Chapter 6 were successful, those are ipso facto arguments for not putting so much weight on theoretical unity. Moreover, while Externalist Accommodationist proposals get us unity at the level of metaphysics, they do this by introducing disunity in their accounts of what's valuable. For instance, Dorsey has to have separate accounts of the existence of reasons and the weight of reasons, and Brink has to say that there can be value both in the objects of choice and in the act of choice. So, everyone has disunity somewhere. It's not clear to me why having it at the level of metaphysics would be worse.

Second, while I argued above that there's no non-question-begging way of articulating the bootstrapping objection to voluntarism, one might still feel the anti-bootstrapping intuition. And if one just finds voluntarism too antecedently implausible, one might be willing to trade some explanatory power in order to avoid commitment to it. This line of reasoning might also lead one to end up in the Brink/Dorsey family of views, given that those can say what hybrid voluntarists

want to say about the normative significance of the will without taking on a voluntarist metaphysics in order to say it.

There is certainly something to this worry. The voluntarist idea is, for many, an initially counterintuitive one. But, it seems to me, the very same things that make it counterintuitive also make it such that it would be really cool if it turns out to be true. Consider again the comparison in Chapter 1 between voluntarism and the divine command theory; hybrid voluntarism says that, in a limited way, we have the kind of normative powers that some have attributed to God. So, before thinking through the bootstrapping objection, it'd be reasonable to treat the surprising nature of the voluntarist claim as evidence that it might give rise to theoretical trouble. But after thinking through the bootstrapping objection and seeing that there's no theoretical trouble forthcoming, what once seemed counterintuitive should now seem intriguing or exciting. Or at least, it seems that way to me, and I invite you to be intrigued as well.

Acknowledgments

In the summer of 2013, Matt Braich and I were at a Padres game, having a beer and chatting about David Brink's paper “The Significance of Desire.” Matt and I wondered what could explain why David leaned so heavily on extensional objections to views that explain an agent's reasons in terms of her desires, but did not find the analogous objections as compelling when raised against views that explain an agent's reasons in terms of her choices. This conversation was the spark of a paper, “These Boots Were Made for Strapping,” in which I defended voluntarism against the bootstrapping objection, but left it open why anyone might want to be a voluntarist. The feedback I got on that paper set me to thinking about the appeal of voluntarism, and this dissertation is the result. So, as a matter of intellectual history, I owe this project to Matt and David. But I owe lots of other people as well, and I owe those two for other things, so a more thorough accounting is in order.

As an undergraduate, when I first entertained the idea of going to graduate school for philosophy, I worried that the idea wouldn't go over well with my family – that it'd strike them as too risky or too self-indulgent. So, when I called home to say I'd decided to apply, it felt wrong to leave the news as a voicemail. Instead, I panicked and said “Mom, call me back, I need to talk about life choices.” It came as a relief to my folks, then, that I only wanted to be a philosopher. At any rate, I was wrong to be worried, so I owe thanks to my mom, Valerie Davia; my step-dad, Warren Luce, and my dad, Robert Davia for supporting that decision and for encouraging the habits of mind that led me to it.

Thanks are also due to the philosophy faculty at Claremont McKenna College, where I

learned how to do philosophy and came to see a career in it as a possibility for me. Particular thanks are due to Alex Rajczi, Amy Kind, Paul Hurley, and Dustin Locke. Alex taught my first philosophy course, and, through how seriously he took his students, provided a model of philosophy as constructive dialogue. I aspire to be as warm and encouraging, and I hope that the influence of his thoroughgoing epistemic humility is evident here. Amy introduced me to the philosophical world outside of value theory. I admire the clarity with which she can frame a debate and her ability to get to the bottom of what a philosophical dispute is really about. I aspire to ask colloquium questions as sharp as hers. Paul got me excited about metaethics, guided me through a senior thesis and the application process for MA programs, and helped me afford to spend the summer in Claremont working on my writing sample by hiring me to index *Beyond Consequentialism*. At every step of the way Paul has been a source of advice and encouragement. For instance: driving me to Los Angeles to attend my first philosophy conference, talking me through decisions about where to send my first journal article, and inviting me back to Claremont for my first colloquium talk. Paul is my model of a kind and supportive colleague. When I returned to CMC to teach as a visiting instructor, my office was right across from Dustin's, and this turned out to be a boon for me. Dustin was a patient and constructive sounding board for teaching ideas, and went well out of his way to make me feel like one of the faculty when imposter syndrome had me feeling like a returning student.

As an MA student at UW Milwaukee, I was lucky to get to spend time with Phil Mack, Chris Mesaros, Emma Friedman, Andrew Flynn, Paul Blaschko, John Hanson, Mark Puestohl, Amy Flowerree, Meica Magnani, Dan Giglio, David Shope, and Stephanie Allen. Thanks to all

of them for philosophically rich conversation and in-joke rich time hanging out at VonTrier, Nessun Dorma, and Ian's Pizza (extra thanks to Chris for encouraging me to try home-brewing with him). Special thanks are due to Michele Palmira for co-authoring what later turned out to be my first publication and to the Justins – Bernstein and D'Ambrosio – for taking me under their wing. I spent years trying to emulate Justin Bernstein as a seminar participant (prepared and serious, but loose and not competitive). I envy Justin D'Ambrosio's healthy skepticism of philosophical consensus, and uncompromising trust in his philosophical instincts. I hope he'd be proud that this dissertation defends a view so many philosophers find implausible.

When I visited UCSD as a prospective student, I asked a few people if they were happy here. One answered with the rhetorical question, “Is *anyone* happy in grad school?” It turns out, though, that I mostly have been, and that's in large part due to the community of graduate students looking out for each other. Thanks to Joyce Havstad, Ben Sheredos, John Dougherty, Theron Pummer, Per Milam, Tanya Hall, Craig Agule, and Amy Berg for welcoming me into that community and teaching me how to be a grad student. Thanks especially to Amy for her commitment to department citizenship. I admire her motivation to do right by her colleagues, and I don't think I would have been as involved or benefited as much from the UCSD community without her example. Special thanks are also due to Craig for generosity with his time, advice, and the sense of play he brings to his work and department life. Thanks also to Tanya for her insights about teaching and willingness to share her process. I'm a much better teacher for having had the opportunity to compare notes with her.

Lots of people have made my time at UCSD rewarding with their philosophical insights

and their friendship. Some of them include: JP Messina, Leo Moauro, Cami Koepke, Ryan Stringer, Henry Argetsinger, Sebastian Speitel, Anncy Thresher, Shawn Wang, Brian Tracz, Emma Duncan, Kathryn Joyce, Bas Tonnison, Danny Weltman, JC Gonzalez, Rebecca Hardesty, Jonathan Knutzen, Will Albuquerque, Dallas Amico, and Richard Vagnino. Though he was only here briefly as a visiting scholar, Robert Wallace deserves a place on this list, and though they're not philosophers by trade, I think of Sally Hargate, Colin Doberstein, Sara Bivin, Sadie Lansdale, and Gabe Kalmuss-Katz as honorary members of the UCSD philosophy community.

When this project was at the prospectus stage, most of the ideas got fleshed out in conversation with Matt Braich, Noel Martin, and Kathleen Connelly. We'd meet weekly to discuss something one of us had written; their feedback (and the regular deadlines) benefited me immensely. Professionally, I'm grateful to Matt for pushing me to think more about how an account like mine could be compatible with what we know from social psychology; my first solo-author publication developed out of trying to respond to objections from him. Personally, I'm grateful to him for all the time we spent talking baseball, shooting pool, and reassuring each other about the various ways grad school can be stressful. In the summer between my first and second years, Matt invited me to join him in an independent study he was doing with Dana Nelkin (more on her shortly). At the time, I was wavering about whether I was good enough at philosophy to hack it here. That he wanted me to be involved helped convince me I was.

Noel and I met as prospective students, and he quickly became one of the people – in philosophy or otherwise – with whom I am at my most unguarded. I admire his sincerity, his sense of whimsy, and his eagerness to stick up for his friends. I'm grateful for all our time spent

cooking together – especially the series of "Wherein Noel and Cory Make Tacos" parties – as well as time riffing, reflecting, talking pedagogy, sharing music and stand up comedy, and catching up over breakfast burritos. I'm grateful, also, for Noel's enthusiasm for karaoke singalongs.

Kathleen has an uncanny ability to reflect your ideas back to you in a way that represents them as far more exciting and important than they were as initially presented. She is also one of the only philosophers I know who, when reading a rough draft, will generate more arguments *for* its conclusion, even if it's one she doesn't antecedently accept. This enthusiasm and generosity extends beyond philosophy to the way she supports and brings out the best in her friends. I'm grateful to be one of them, having benefited especially from her sympathetic ear and quickness to laugh.

Though not in our writing group, Rosalind Chaplin also deserves special thanks. Lots of the ideas here have gotten clearer for having been bounced off of Roz. She is wonderful as a coffee-break companion, as a billiards partner, as a colloquium-question-workshopper, and for helping to figure out what to say to Reviewer #2. I'm lucky that we made the transition from work friends to friends full stop, and I'm cautiously optimistic that her latent Cardinals fandom can be rekindled into full-blown baseball interest.

This project benefited from a dedicated and generous supervising committee. I've already mentioned one way in which that committee's chair, David Brink, led me to the ideas pursued here, but beyond the causal story I admire a lot about how David does philosophy. In particular, David has an incredible ability to see the whole board, cleanly mapping the conceptual space and

which moves are available to whom. I'm also impressed with (and a bit intimidated by) the way his work sympathetically engages the history of philosophy, finding new and plausible approaches where one might not have expected them. Although this project didn't turn out to have the historical component I once thought it might, I hope David's resourcefulness and systematicity are represented here. He also has a great dog (and a newer, less great but still pretty good dog).

I'm indebted to Dana Nelkin for seeing the best in this project before I knew quite what I was up to. Whenever I presented Dana with a half-baked idea, she would talk it through with me and point me toward resources that helped lead me to the fully-baked version. I'm grateful for that, for her encouragement, and for her willingness to read things with me no matter how disconnected from her own projects. I admire Dana's ability to see connections between what might otherwise look like disparate literatures and the sympathy with which she reconstructs opposing views. I hope I've emulated that to some extent here.

I worked most closely with David and Dana, but the influence of the rest of the committee shows up all over this project. Talking with Richard Arneson helped me identify and clarify my metaphilosophical assumptions and motivations. Clinton Tolley's questions at my prospectus defense about why we should think of voluntarism as interestingly distinct from internalism drove both the framing choices I made in Chapter 1 and some of the arguments I ended up making in Chapter 6. Craig McKenzie's psychology seminar on human rationality directly led to my discussion of the St. Petersburg Paradox in Chapter 5 and framed a lot of what I think about the role of reflection in agency. Though not officially on the committee, Monique

Wonderly nonetheless helped me clarify a lot of argumentative moves – especially in Chapters 2 and 4 – and to organize my thinking in Chapter 6. Thanks also to Manuel Vargas for stepping in to join the committee at the last minute, thereby enabling me to solve a defense scheduling paradox.

One more philosopher deserves special mention. I'm grateful to Ruth Chang for making it possible for me to be a visiting student at Rutgers, and for being so kind, welcoming, and generous with her time and feedback while I was there. Conversations with her were where the reply to the "Externalist Accommodationist" first started to take shape; those are now some of the arguments of which I'm most proud. Ruth's work has a dazzling combination of analytical rigor, big-picture wonder, and attention to everyday experience. After having been inspired by it from afar, it was a thrill to get to do philosophy with her.

The most important person to thank, though, is my partner, Sarah Nathan. Her willingness to work hard for important goals has been a model for me, her zest for life keeps me fresh, her company makes my days meaningful. I couldn't have done this without her.

Works Cited

Alvarez, Maria. (2010) *Kinds of Reasons*. Oxford: Oxford University Press.

Arneson, Richard (1999) "Human Flourishing Versus Desire Satisfaction," *Social Philosophy and Policy* 16 (1). pp. 113-142.

Balaguer, Mark (2010) *Free Will as an Open Scientific Problem*. Cambridge: The MIT Press.

Behrends, Jeff (2015) "Problems and Solutions for a Hybrid Approach to Grounding Practical Normativity," *Canadian Journal of Philosophy* 45 (2). pp. 159-178.

Bentham, Jeremy (1789) *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press.

Blackburn, Simon (1993) *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Bratman, Michael (2018) *Planning, Time, and Self-Governance: Essays in Practical Rationality.* Oxford: Oxford University Press.

Brandt, Richard (1979) *A Theory of the Good and the Right*. Amherst: Prometheus Books.

Brink, David (1989) *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Brink, David (1997a) "Moral Motivation," *Ethics* 108 (1). pp. 4-32.

Brink, David (1997b) "Self-Love and Altruism," *Social Philosophy and Policy* 14 (1). pp. 122-157.

Brink, David (2003) *Perfectionism and the Common Good: Themes in T.H. Green*. Oxford: Clarendon Press.

Brink, David (2006) "The Autonomy of Ethics" in *The Cambridge Companion to Atheism* ed. M. Martin. Cambridge: Cambridge University Press. pp. 149-165.

Brink, David (2008) "The Significance of Desire" in *Oxford Studies in Metaethics: Volume 3* ed. R. Shafer-Landau. Oxford: Oxford University Press. pp. 5-46.

Brink, David (2014) "Principles and Intuitions in Ethics: Historical and Contemporary Perspectives," *Ethics* 124 (4). pp665-694.

Broome, John (2018) “Reasons Fundamentalism and What Is Wrong With It” in *The Oxford Handbook of Reasons and Normativity* ed. D. Star. Oxford: Oxford University Press. pp. 297-320.

Callahan, Laura (ms) “Grounding the Normativity of Epistemic Rationality: Responsivism”

Chakravartty, Anjan (2017) *Scientific Ontology: Integrating Naturalized Metaphysics and Voluntarist Epistemology*. Oxford: Oxford University Press.

Chang, Ruth (2002) “The Possibility of Parity,” *Ethics* 112 (4). pp. 659-688.

Chang, Ruth (2009) “Voluntarist Reasons and the Sources of Normativity” in *Reasons for Action* eds. D. Sobel and S. Wall. Cambridge: Cambridge University Press. pp 243-171.

Chang, Ruth (2012) “Are Hard Choices Cases of Incomparability?” *Philosophical Issues* 22 (1). pp. 106-126

Chang, Ruth (2013a) “Commitment, Reasons, and the Will” in *Oxford Studies in Metaethics: Volume 8* ed. R. Shafer-Landau. Oxford: Oxford University Press. pp. 74-113.

Chang, Ruth (2013b) “Grounding Practical Normativity: Going Hybrid,” *Philosophical Studies* 164 (1). pp. 163-187.

Coates, Jonathan (2017) “A Wholehearted Defense of Ambivalence,” *The Journal of Ethics* 21 (4). pp. 419-444.

Cohen, G.A. (2014) “Reason, Humanity, and the Moral Law” in *Lectures on the History of Moral and Political Philosophy* ed. J. Wolff. Princeton: Princeton University Press. pp 305-324.

Connelly, Kathleen (ms) “Accountability and Unconscious Reasons”

Copp, David (2003) “Why Naturalism?” *Ethical Theory and Moral Practice* 6 (2). pp. 179-200.

Cueno, Terence (2007) “Recent Faces of Moral Nonnaturalism,” *Philosophy Compass* 2 (6). pp. 85-879.

Dancy, Jonathan (2000) *Practical Reality*. Oxford: Oxford University Press.

Dancy, Jonathan (2004) *Ethics Without Principles*. Oxford: Oxford University Press.

Davia, Cory (2018) “Reflection Without Regress,” *Pacific Philosophical Quarterly*. Published online, print version forthcoming.

Davidson, Donald (1963) “Actions, Reasons, and Causes,” *The Journal of Philosophy* 60 (23). pp. 685-700.

Dreier, Jamie (1997) “Humean Doubts about the Practical Justification of Morality” in *Ethics and Practical Reason* eds. G. Cullity and B. Gaut. Oxford: Oxford University Press. pp. 81-100.

Dreier, Jamie (2015) “Can Reasons Fundamentalism Answer the Normative Question?” in *Motivational Internalism* eds. G. Bjornsson, C. Strandberg, R. Olinder, J. Eriksson, and F. Bjorklund. Oxford: Oxford University Press. pp 167-181.

Dorsey, Dale (2016) “The Normative Significance of Self,” *Journal of Ethics and Social Philosophy* 10 (1). pp. 1-25.

Dworkin, Ronald (1986) *Law's Empire*. Harvard: Harvard University Press.

Enoch, David (2006) “Agency, Shmagency: Why Normativity Won't Come From What is Constitutive of Action,” *Philosophical Review* 115 (2). pp. 169-198.

Enoch, David (2011a) *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

Enoch, David (2011b) “Shmagency Revisited” in *New Waves in Metaethics* ed. M. Brady. Palgrave-Macmillan.

Ferrero, Luca (2009a) “Constitutivism and the Inescapability of Agency” in *Oxford Studies in Metaethics: Volume Four* ed. R. Shafer-Landau. Oxford: Oxford University press. pp. 303-333.

Ferrero, Luca (2009b) “What Good is a Diachronic Will?” *Philosophical Studies* 144 (3). pp. 403-430.

Finlay, Stephen (2009) “The Obscurity of Internal Reasons,” *Philosophers' Imprint* 9 (1). pp. 1-22.

Fogal, Daniel (2016) “Reasons, Reason, and Context” in *Weighing Reasons* eds. E. Lord and B. Maquire. Oxford: Oxford University Press. pp. 74-103.

Foot, Philippa (1972) “Morality as a System of Hypothetical Imperatives,” *Philosophical Review* 81 (3). pp. 305-316.

Frankfurt, Harry (1971) “Freedom of the Will and the Concept of a Person,” *Journal of Philosophy* 68 (1). pp. 5-20.

Frankfurt, Harry (2004) *The Reasons of Love*. Princeton: Princeton University Press.

Gibbard (1990) *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press.

Gibbard, Allan (1999) "Morality as Consistency in Living: Korsgaard's Kantian Lectures," *Ethics* 110 (1), pp. 140-164.

Gibbard, Allan (2003) *Thinking How to Live*. Cambridge: Harvard University Press.

Green, T.H. (1883) *Prolegomena to Ethics*. Oxford: Clarendon Press.

Grice, H.P. (1975) "Logic and Conversation" in *Speech Acts* eds. P. Cole and J.L. Morgan. New York: Academic Press. pp 41-58.

Grube, G.M.A. (2002) trans. Plato. *Five Dialogues*. Indianapolis: Hackett Publishing Company.

Gunnarsson, Logi (2014) "In Defense of Ambivalence and Alienation," *Ethical Theory and Moral Practice* 17 (1). pp. 13-26.

Harman, Gilbert (1977) *The Nature of Morality*. New York: Oxford University Press.

Herman, Barbara (2007) *Moral Literacy*. Harvard: Harvard University Press.

Hobbes, Thomas. (1651) *Leviathan*. London: Andrew Crooke.

Hurka, Thomas (1993) *Perfectionism*. Oxford: Oxford University Press.

Hurley, Paul (2018) "Consequentialism and the Standard Story of Action," *The Journal of Ethics* 22 (1). pp. 25-44.

Hume, David (1738) *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning Into Moral Subjects*. London: John Noon.

Irwin, Terence (2008) *The Development of Ethics, Volume Two: From Suarez to Rousseau*. Oxford: Oxford University Press.

Kane, Robert (2012) "Torn Decisions, Luck, and Libertarian Free Will: Comments on Balaguer's Free Will as an Open Scientific Problem," *Philosophical Studies* (1). pp 1-8.

Kant, Immanuel (1785) *Groundwork of the Metaphysics of Morals*. Berlin: De Gruyter.

Kelly, Thomas & McGrath, Sarah (2010) *Philosophical Perspectives* 24 (1). pp. 325-359.

Kornblith, Hilary (2012) *On Reflection*. Oxford: Oxford University Press.

Korsgaard, Christine (1996) *The Sources of Normativity.* Cambridge: Cambridge University Press.

Korsgaard, Christine (1997) "The Normativity of Instrumental Reason" in *Ethics and Practical Reason* eds. G. Cullity and B. Gaut. Oxford: Oxford University Press. pp. 215-254.

Korsgaard, Christine (2009) *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Leopold, Aldo (1972) *Round River: From the Journals of Aldo Leopold*. Oxford: Oxford University Press.

Locke, Dustin (2014) "Darwinian Normative Skepticism" in *Challenges to Moral and Religious Belief: Disagreement and Evolution* (eds.) M. Bergmann and P. Kain. Oxford: Oxford University Press.

Mackie, J.L. (1977) *Ethics, Inventing Right and Wrong*. Harmondsworth: Penguin.

Manne, Kate (2016) "Democratizing Humeanism" in *Weighing Reasons* eds. E. Lord and B. Maquire. Oxford: Oxford University Press. pp. 123-140

Manne, Kate (2014) "Interalism about Reasons: Sad but True?" *Philosophical Studies* 167. pp89-117.

Markovits, Julia (2014) *Moral Reason*. Oxford: Oxford University Press.

McDowell, John (1978) "Are Moral Requirements Hypothetical Imperatives?" *Proceedings of the Aristotelian Society* 52 (1). pp. 13-42.

Mill, J.S. (1863) *Utilitarianism*. London: Parker, Son, and Bourn, West Strand.

Moss, Sarah (2018) "Moral Encroachment," *Proceedings of the Aristotelian Society* 118 (2). pp. 177-205.

Nagel, Thomas (1970) *The Possibility of Altruism*. Oxford: Oxford Clarendon Press.

Nagel, Thomas (1986) *The View from Nowhere*. Oxford: Oxford University Press.

Owens, David (2012) *Shaping the Normative Landscape*. Oxford: Oxford University Press.

Parfit, Derek (2011) *On What Matters: Volume One*. Oxford: Oxford University Press.

Parfit, Derek (2017) *On What Matters: Volume Three*. Oxford: Oxford University Press.

Paul, L.A. (2015) *Transformative Experience*. Oxford: Oxford University Press.

Perry, John (1976) "The Importance of Being Identical" in *The Identities of Persons* ed. A. Rorty. Oakland: University of California Press. pp. 67-90.

Pufendorf, Samuel (1672) *Of the Law and Nature of Nations*. Lund: Junghans.

Railton, Peter (1986) "Facts and Values," *Philosophical Topics* 14 (2). pp. 5-31.

Railton, Peter (1992) "Pluralism, Determinacy, and Dilemma," *Ethics* 102 (4). pp. 720-742.

Railton, Peter (1997) "On the Hypothetical and Non-hypothetical in Reasoning about Belief and Action" in *Ethics and Practical Reason* eds. G. Cullity and B. Gaut. Oxford: Oxford University Press. pp. 53-79.

Railton, Peter (2014) "The Affective Dog and Its Rational Tale: Intuition and Attunement," *Ethics* 124 (4). pp. 813-859.

Rajczi, Alex (2011) "The Argument from Self-Creation: A Refutation of Act-Consequentialism and a Defense of Moral Options," *American Philosophical Quarterly* 48 (4). pp. 315-332.

Rawls, John (1971) *A Theory of Justice*. Cambridge: Harvard University Press.

Raz, Joseph (1986) *The Morality of Freedom*. Oxford: Oxford University Press.

Ridge, Michael (2007) "Anti-reductionism and Supervenience," *Journal of Moral Philosophy* 4 (3). pp. 330-348.

Rosati, Connie (1995a) "Persons, Perspectives, and Full Information Accounts of the Good," *Ethics* 105 (2). pp. 296-325.

Rosati, Connie (1995b) "Naturalism, Normativity, and the Open Question A rgument," *Nous* 29 (1). pp. 46-70.

Scanlon, T.M. (2013) *Being Realistic about Reasons*. Oxford: Oxford University Press.

Scanlon, T.M. (1998) *What We Owe to Each Other.* Cambridge: Belknap Press.

Schroeder, Mark (2005) "Realism and Reduction: The Quest for Robustness," *Philosophers' Imprint* 5 (1). pp. 1-18.

Schroeder, Mark (2007a) "Reasons and Agent-Neutrality," *Philosophical Studies* 135 (2). pp. 279-306.

Schroeder, Mark (2007b) *Slaves of the Passions*. Oxford: Oxford University Press.

Schroeder, Mark (2008) "Having Reasons," *Philosophical Studies* 139 (1). pp. 57-71.

Schroeder, Mark (2018) "Getting Perspective on Objective Reasons," *Ethics* 128 (2). pp. 289-319.

Setiya, Kieran (2017) *Midlife: A Philosophical Guide*. Princeton: Princeton University Press.

Shafer-Landau, Russ (2003) *Moral Realism: A Defence*. Oxford: Oxford University Press.

Smith, Michael (1994) *The Moral Problem*. Oxford: Blackwell Publishing.

Smithm Matthew (2016) "One Dogma of Philosophy of Action," *Philosophical Studies* 173 (8). pp. 2249-2266.

Street, Sharon (2006) "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127 (1). pp. 109-166.

Street, Sharon (2009) "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters," *Philosophical Issues* 19 (1). pp. 273-298.

Stringer, Ryan (2018) "Realist Ethical Naturalism for Ethical Non-naturalists," *Philosophical Studies* 175 (2). pp. 339-362.

Sturgeon, Nicholas (2007) "Ethical Naturalism" in *The Oxford Handbook of Ethical Theory* ed. D. Copp. New York: Oxford University Press. pp. 91-121.

Sussman, David (2015) "Morality, Self-Constitution, and the Limits of Integrity" in *Why Be Moral?* eds. R. Louden and B. Himmelmann. Berlin: De Gruyter. pp. 123-140.

Titelbaum, Michael (2019) "Reason Without Reasons For" in *Oxford Studies in Metaethics: Volume 14* ed. R. Shafer-Landau. Oxford: Oxford University press. pp. 189-215.

van Roojen, Mark (2017) “Review of Derek Parfit, *On What Matters Volume Three*” in *Notre Dame Philosophical Reviews*.

Velleman, J.D. (1997) “Deciding How to Decide,” in *Ethics and Practical Reason* eds. G. Cullity and B. Gaut. Oxford: Oxford University Press. pp. 29-52.

Velleman, J.D. (2000) *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Velleman, J.D. (2009) *How We Get Along*. Cambridge: Cambridge University Press.

Velleman, J.D. (2013) *Foundations for Moral Relativism*. Cambridge: Open Book Publishers.

Watson, Gary (2009) “Promises, Reasons, and Normative Powers” in *Reasons for Action* eds. D. Sobel and S. Wall. Cambridge: Cambridge University Press. pp. 155-178.

Wedgwood, Ralph (2007) *The Nature of Normativity.* Oxford: Oxford University Press.

Wedgwood, Ralph (2015) “The Pitfalls of 'Reasons',” *Philosophical Issues* 25 (1). pp. 123-143.

Weisberg, Jonathan (2012) “The Bootstrapping Problem,” *Philosophy Compass* 7 (9). pp. 597-610.

Westlund, Andrea (2013) “Deference as a Normative Power,” *Philosophical Studies* 166 (3). pp. 455-474.

Whiting, Jennifer (1986) “Friends and Future Selves,” *Philosophical Review* 95 (4). pp. 547-80.

Williams, Bernard (1973) “A Critique of Utilitarianism” in *Utilitarianism: For and Against* eds. J. Smart and B. Williams. Cambridge: Cambridge University Press. pp. 77-150.

Williams, Bernard (1981) “Internal and External Reasons” in *Rational Action* ed. R. Harrison. Cambridge: Cambridge University Press. pp. 101-113.

Williams, Bernard (1989) “Internal Reasons and the Obscurity of Blame” in *Logos Volume 10* ed. W. Prior. Santa Clara University. pp. 1-11.

Wodak, Daniel (ms) “Who's On First?” forthcoming in *Oxford Studies in Metaethics: Volume 15* ed. R. Shafer-Landau. Oxford: Oxford University Press.