TOWARDS THE AUTONOMY OF ETHICS: SKEPTICISM, AGENCY, AND NORMATIVE COMMITMENT

by

Hille Paakkunainen

MA, University of Glasgow, 2003

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

Faculty of Arts and Sciences

This dissertation was presented

by

Hille Paakkunainen

It was defended on

July 19, 2011

and approved by

John McDowell, Distinguished University Professor of Philosophy

Stephen Engstrom, Professor of Philosophy

Peter Machamer, Professor, History and Philosophy of Science

Dissertation Director: Kieran Setiya, Associate Professor of Philosophy

TOWARDS THE AUTONOMY OF ETHICS: SKEPTICISM, AGENCY, AND NORMATIVE COMMITMENT

Hille Paakkunainen, PhD

University of Pittsburgh, 2011

How may we try to answer the central question of ethics, the question how one should live? Understood as concerning the good lives of rational agents *qua* rational, the question concerns the standards of practical reason. How may we vindicate a view about those standards—an *ethical view*, for short?

This dissertation examines whether it is possible to vindicate an ethical view without begging any first-order normative questions against skeptics in the process. I argue that it is not. If there are sound arguments for ethical views, they must rely on premises that, while true, beg some first-order normative question against a possible skeptic. I call this thesis *the autonomy of ethics*. The result is that sound ethical argumentation is disturbingly partisan: sound arguments in ethics cannot be seen to be sound by anyone who does not already share the right first-order view to at least some extent.

I argue for the autonomy of ethics by examining attempts to avoid it. *Constitutivism* seeks to ground ethics in the metaphysics of agency. *Metasemantic* strategies seek to ground ethics in the conditions of concept-possession, and in the implicit normative commitments that such conditions purportedly involve. Closely related *metapragmatic* strategies seek to ground ethics in the conditions of using concepts in judgments or in reasoning. Against each strategy, I argue that the relevant conditions—the conditions of agency, of concept-possession, and of concept use—are normatively neutral. I further argue that, given the failure of these strategies, there is no further way to avoid the autonomy of ethics. The only possible sound arguments in favor of ethical views are ethically partisan in the way outlined.

One way of putting this conclusion is that there is no purely metaethical way of vindicating any ethical view. If there can nonetheless be objective truths in ethics, their possibility cannot depend on their having a purely metaethical grounding.

TABLE OF CONTENTS

PREFACE

I have been incredibly lucky to study philosophy at Pittsburgh for these past eight years. I am grateful, most of all, to my advisor Kieran Setiya. Kieran is simply a terrific dissertation advisor. He has consistently held me to high standards, and expectations matter. If I had to pick one thing that I most hope to have learned from him, it would be to have internalized some of those high standards. His own work is an example of philosophical rigor and integrity, and I owe him a tremendous debt of gratitude for setting that example, and for patiently teaching me how to approximate to its standards, however imperfectly I may do so. I would not have been able to write this dissertation without his encouragement, example, guidance, and acute criticisms. I will miss working with him.

I am also very grateful to John McDowell, whose work and philosophical presence at Pittsburgh have influenced the shape of this dissertation profoundly, yet in ways that I find difficult to catalogue. I always felt that if I needed to get my head screwed back on right, I only needed to go talk to him. Yet his work, and my discussions with him, always also left me with the most persistently nagging questions. I am certain that some of this is due to my failure to see things clearly. But I believe that much of it is also because his work hews close to bedrock—to questions that nag not because they are based on a confusion we still have, but because they are real problems of life for beings like us. I have not yet found a way to deal with those problems. But the overall shape of my dissertation is an attempt to (re)articulate one such problem.

Stephen Engstrom's presence in my dissertation committee has been very important to me. I absolutely love talking to him about philosophy. His challenging comments have always improved my work greatly. And it was when talking to Steve about Kant's conception of the good will about two and a

half years ago that I first began to feel really comfortable taking my time to think through things in the presence of others. His calm and measured way of discussing philosophy has given me the confidence and courage to try to formulate my own thoughts precisely and carefully at my own pace, and to resist the rushed atmosphere that intellectual exchanges in a high-powered graduate school can sometimes take on. Steve has taught me a great deal about careful thinking, and I am immensely grateful to him for that.

I also owe a debt of gratitude for Peter Machamer for being my external reader. One of my greatest regrets about my dissertation work is that I did not take enough advantage of his presence at Pittsburgh. On the occasions on which we did speak, his insights and feedback were always fresh and quite different from the feedback I received from anyone else. Many of the criticisms he has raised will be topics for future work for me.

I would like to thank my friends and loved ones for all their support and encouragement over the years, and for keeping me sane. Kim Frost has been my best friend, philosophically and otherwise, for many years now. He has gracefully listened to, read, and helped me to improve, half-baked versions of almost all of the ideas and arguments in the dissertation, many times over. I always have the best and most deeply gratifying philosophical discussions with him. I continue to have the good fortune to learn from his work, and from his example as a human being.

Evgenia Mylonaki and James Pearson have both left an indelible mark on me as a philosopher and a person. Evgenia's is a unique and infectious brand of humanism, and she is intellectually honest to the core. I admire her and her work very much. James somehow manages to combine enviable level-headedness with utter wackiness. Our conversations about constitutivism and skepticism in particular have been exceptionally important to me; and learning from his work has helped me to make many historical and philosophical connections I would otherwise have missed. I look forward to a lifelong philosophical friendship with both.

I am also grateful to Tim Willenken, Greg Strom, and Sasha Newton, each of whom has been a highly positive influence for me, especially in my early years at Pitt; to Dan Addison, for his generous good humor and enduring friendship, and for trying to teach me what Kant means by freedom; to Kathryn

Lindeman, especially for our many conversations about constitutivism and about the nature of normativity; to Tyke Nunez, for his philosophical poise and insight as an interlocutor; and to Karl Schafer, whose sharpness and humor I admire, and whose comments on my second chapter and on portions of the third greatly improved them.

My dear friend Stacy Hoffman told me, when I was agonizing over how to make my preface neither too sappy nor too snappy, that it was my chance to be sappy. I trust her judgment in many things, and here too I have erred on the side of her suggestion. I am grateful to her and to Michael Cuccaro for many years of friendship. They are both some of the kindest and most generous people I know, and have played a very large part in making Pittsburgh a home for me.

Lastly, I would like to thank my family in Finland, who have always been very supportive, even from afar, and whom I would like to see so much more often. I dedicate this dissertation to them.

* * *

Parts of this dissertation were written with the support of Jenny and Antti Wihuri Foundation. I am very grateful for the funding. I am also grateful to audiences at Pittsburgh, UCLA, Dartmouth and Syracuse for helpful and stimulating discussions of portions of the first three chapters.

INTRODUCTION

How may we try to answer the central question of ethics, the question how one should live? Understood as concerning the good lives of rational agents *qua* rational, the question concerns the standards of practical reason. How may we try to vindicate a view about those standards—an *ethical view*, for short?

One type of vindication proceeds in first-order normative terms, making essential appeal, in the course of the vindication, to parts of the very view being vindicated. Such arguments may be sound if their premises are true, but the procedure can seem objectionably question-begging. For such an argument cannot be seen to be sound by anyone who does not already share the right first-order view to at least some extent. Unless there is some reason why we are all bound to have the right view, arguments proceeding in first-order terms are disturbingly partisan. They leave us with no intellectual response to skepticism besides rhetoric: even if sound, they cannot shore up the confidence of anyone, including ourselves, who might be genuinely wondering about the credentials of an ethical view.

I call the thesis that the only possible sound arguments in favor of any ethical view are themselves ethically partisan in this way *the autonomy of ethics*. My dissertation argues for the autonomy of ethics by examining attempts to avoid it.

We could transcend partisanship if we could argue that there are standards of practical reason to which we are all necessarily committed, however implicitly: such standards would be skeptic-proof. *Constitutivism* tries to find such implicit normative commitments in the metaphysics of agency. *Metasemantic* strategies try to find such implicit normative commitments in the conditions of possessing the concepts needed to so much as formulate, let alone doubt, a given ethical view. Closely related *metapragmatic* strategies attempt to find such implicit normative commitments in the conditions of using

ethically relevant terms and concepts in judgments or in reasoning. In fact, as I argue, these strategies look particularly powerful because they look capable of showing not only that certain ethical views are indubitable, but also that they are *true*: the conditions of agency, of concept-possession, and of concept use, each promise in their different ways to provide an objective grounding for ethics.

Against each strategy, however, I argue that the relevant conditions—conditions of agency, of concept-possession, and of concept use—are normatively neutral. These conditions cannot be used to argue for any ethical view over another. I further argue that, given the failure of these strategies, there is no further possible way of avoiding the autonomy of ethics. The only possible sound arguments in favor of ethical views are ethically partisan: the autonomy of ethics is true. One way of putting this conclusion is that there is no purely metaethical way of vindicating any ethical view. Even if some ethical view is true, it cannot be shown to be so by providing that view a purely metaethical grounding. Whether there can nonetheless be objective truths in ethics depends upon whether such objectivity is possible without metaethical grounding.

The dissertation comprises four chapters, largely self-standing, but tied together by the theme of examining attempts to avoid the autonomy of ethics. The first three chapters concern constitutivism, which attempts to ground the standards of practical reason, or as call I them, 'practical norms', in the metaphysics of agency. The fourth concerns metasemantic and metapragmatic strategies, which attempt to ground standards of practical reason in the conditions of concept-possession and of concept use, respectively.

Chapter 1: Constitutivism in practical reason

Constitutivists attempt to derive "ought"-claims, about the standards of practical reason, from "is"-claims, about the nature of agency. Chapter 1 examines how such derivations are supposed to work. I argue that, although a seemingly powerful "function" argument doesn't help to explain the derivation, there is a

different argument that does. According to that argument, given a specific sort of "is"-claim about agency, the constitutivist derivation of "ought" from "is" is unassailable. Furthermore, in defending "constitutive" practical norms, this argument looks to leave no room for any non-constitutive practical norms at all. The result is that anyone wondering what the practical norms governing rational agency are has only two options: either embrace constitutivism, or challenge the specific sort of "is"-claim on which the constitutivist argument relies. Either way, I argue, we must seriously engage in-depth theses in the philosophy of action. Philosophy of action is inescapable for metaethics.

This chapter motivates a serious look at the nature of agency, with a view to evaluating the prospects for constitutivism. The aim of the next two chapters is then to challenge the sorts of "is"-claim about the nature of agency that would validate the constitutivist derivation of "ought" from "is." I divide my investigation of agency into two parts, one focusing on the notion of *efficacy* in agency (ch.2), the other on the notion of *autonomy* (ch.3). I take chapters 2 and 3 together to amount to a defense of a conception of agency, of its efficacy, and of its autonomy, that is comprehensively anti-constitutivist. The result is that, if there is going to be a normatively non-question-begging grounding for practical norms, it has to be found elsewhere than in the nature of agency.

Chapter 2: Instrumentalism and practical rule-following

Some recent defenders of instrumentalism, the view that good practical reasoning and reasons are instrumental in form, have attempted to support their view by constitutivist appeals to the nature of agency. We can view all such constitutivist arguments for instrumentalism as stemming from one central intuition, elaborated in different ways: that agency is essentially *efficacious*, and that to be efficacious, one must reason instrumentally and act for instrumental reasons. Against constitutivist defenses of instrumental norms, chapter 2 argues that the notion of efficacious agency is normatively non-committal. I analyze agency as a matter of practical rule-following, of reasoning in accord with practical rules; and I argue that efficacious practical rule-following is not necessarily a matter of following an instrumental rule

of reasoning in particular. Indeed, the nature of efficacy in practical rule-following *cannot* be accounted for by the idea of following an instrumental rule of reasoning. There are two reasons for this. First, there are cases of acting "from" character dispositions such as friendship or courage that are instances of efficacious practical rule-following but elude instrumentalist analysis. And second, the idea that efficacious practical rule-following could be explained by appeal to following an instrumental rule leads to a rule-following regress.

My argument does not deny that practical reasoning can, and often does, have means and ends as part of its *topic*. But I show how having means and ends as a topic is compatible with practical reasoning whose form is non-instrumental. As a result, a norm enjoining instrumental reasoning is not a "constitutive norm" of agency: we cannot derive the authority of instrumental reason from the conditions of agency.

Chapter 3: Autonomy and contingency

Rational agency seems to be essentially autonomous in some sense. Agents are not just passive bystanders to their bodily movements, but rather self-determining authors of their actions. In chapter 3, I attempt to understand what truth there is in such pronouncements. I show why the issue matters, connecting it to the assessment of constitutivism. And I argue that the sense in which agency is essentially autonomous does not support any standards of practical reason, contrary to some constitutivists' claims. Constitutivists who wish to lean on the notion of autonomy in agency to vindicate a view about the standards of practical reason must hold that there is some disposition of practical reasoning that is metaphysically necessary for autonomous rational agency as such. I argue, to the contrary, that autonomous rational agency can be exercised through dispositions that are *contingent* to the nature of autonomous agency as such. As a result, disputes concerning the standards of practical reason cannot be settled by constitutivist appeals to the nature of autonomous agency. The argument proceeds through examining the central concepts in terms of which autonomy is often explicated by constitutivists

themselves: self-determination, psychic unity, identification, universality, self-legislation, self-knowledge, and self-understanding. In each case, I either account for the concept in non-constitutivist terms, or reject its centrality to an account of autonomy.

Chapter 4: Is there a metasemantic route to ethical truth?

Aside from constitutivist derivations of "ought" from "is," what other options are there for attempting to avoid the autonomy of ethics? Some recent work in epistemology has attempted to ground objective norms of theoretical reasoning in *metasemantic* facts, about what it is to possess certain concepts or to understand the meanings of certain terms. Through interrogating some of this work, I examine, in chapter 4, how a metasemantic derivation of "ought" from "is" might go in the case of norms of practical reasoning. I develop two different metasemantic argument schemas; and I further identify two closely related argument schemas we might call *metapragmatic*, appealing to the conditions of concept *use*. I evaluate these arguments in the context of both thick and thin normative concepts, in each case arguing that their key premises are either false, normatively question-begging, or equivocal. In seeing where the arguments go wrong, we see quite generally why we can both possess and use normative concepts, whether thick or thin, without thereby undertaking any implicit normative commitments of the sort that could help vindicate specific practical norms.

The conclusion of this final chapter is also the conclusion of the dissertation. It shows how the failure of constitutivist, metasemantic, and metapragmatic strategies together entails the autonomy of ethics; and it briefly considers the consequences for objectivity in ethics.

1.0 CONSTITUTIVISM IN PRACTICAL REASON

1.1 INTRODUCTION

If there are practical norms—if there are truths about how we ought to live, act and deliberate—what grounds their authority, and what gives them their specific content? The possibility of serious-minded skeptical questioning of particular normative claims helps to make this question gripping. It seems that we should have something to say, to someone skeptical of the norms we espouse, about why those putative norms, as opposed to some different ones, are truly authoritative. In particular, it seems that our justification of a specific set of practical norms should not rely on further practical norms that the skeptic could in turn doubt: to answer the skeptic, we should be able to explain, in principled and normatively non-question-begging terms, why the specific norms we espouse are authoritative. (Any actual skeptic may of course remain unconvinced by the explanation; but the point is that this should not be the fault of the explanation.[1]) Otherwise, it may seem that our specific normative claims are in principle no different from heartfelt but rationally arbitrary declarations of allegiance to a way of life, our own confidence in them backed up at best by persuasive rhetoric.[2]

---

[1] That is, what matters is not that the explanation is dialectically effective against actual skeptics, but just that the explanation is a sound argument that does not beg any normative questions against the skeptic.

[2] Embracing this seeming rational arbitrariness may take several forms: dogmatism, pragmatism, relativism, nihilism. If the skeptic *cannot* be satisfactorily answered, however, we should ask to what extent we are really stuck with rational arbitrariness. In particular, we should ask whether a conception of rational justification that inescapably lands us in a putatively unjustifiable (and in this sense rationally arbitrary) position is mistaken. I touch upon this at the very end of chapter 4.

Constitutivists in practical reason have tried to answer this demand for a skeptic-proof grounding of practical norms by appeal to the nature of agency.[3] Agency, on this view, is metaphysically fundamental with respect to norms, in the sense that practical norms are already implicit in agency: whenever agents exist, practical norms do, and indeed, the practical norms that exist do so *because* agents exist. Practical norms are "constitutive norms" of agency. The constitutivist seeks to give an account of agency that articulates how this is so.[4]

In giving such an account, the constitutivist seeks to answer three questions at once. Aside from the question what agency is, and the question what grounds practical norms, she also gives an account of the content of the domain of practical norms. That is, she offers an account of what good deliberation and good reasons for action amount to—an account, in short, of how rational agents, just as such, should live. The constitutivist strategy, then, promises to be very powerful. In explaining the nature of agency, how agency grounds the authority of practical norms, and the content of those norms, she addresses three of the most pressing questions in practical philosophy—and all this without begging normative questions against the skeptic.[5]

Constitutivism can seem *prima facie* odd. Why think that what we are like—even what we are essentially like, *qua* agents capable of deliberating and acting for reasons—corresponds to, let alone explains, how we *should* live, deliberate and act? A similar oddity is shared by any metaphysics of norms that seeks to ground normative truths in seemingly non-normative truths. As Hume observed, the transition from the "copulations of propositions, *is*, and *is not*" to propositions expressing a "new relation of affirmation," namely, "*ought*, or *ought not*," requires attention and explanation (*Treatise* 3.1.2:27).[6]

---

[3] E.g. Dreier 1997, Korsgaard 1996, 2008b, 2009, Vogler 2002, Velleman 2000, 2009, Smith 2009, 2010.
[4] The conception of metaphysics made use of here is well explicated in Schaffer 2009.
[5] Although the constitutivist seeks to answer many questions at once, one question she need not be in the business of answering is how agents ordinarily know what the authoritative practical norms are. The constitutivist derivation of practical norms from the nature of agency need not be agents' normal epistemological route to knowledge of these norms. Agents might know how they should act and reason without familiarity with the constitutivist's argument.
[6] Though the "ought"-claims Hume is concerned with in the cited passage are specifically those of morality.

Some recent constitutivists have sought the relevant explanation in the idea, perhaps most familiar from the ethics of Plato and Aristotle, of *excellent functioning*. Christine Korsgaard makes this strategy most explicit, and defends constitutivism through it, in her (2009) book *Self-Constitution* and in related material; but a similar strategy can be found, for instance, in some of Michael Smith's recent work (2010). According to this strategy, agency has a function—a characteristic or defining activity—that in and of itself entails a view about what good deliberation and good reasons for action amount to. This is because a function just is the sort of thing that can be performed either well or badly, excellently or defectively. Performing the function of agency is just what it is to be an agent: this is the "is"-claim. And performing one's function *well* or *excellently* or *perfectly* is just what it is to be good *qua* agent: one does exactly what agents as such ought to do, and how they ought to do it. This is the "ought"-claim. If the "function" of agency just specifies the nature of agency, the activity most essential to it; and if performance of the function admits of degrees of better or worse performance; then it may seem as though we can derive, from claims about the nature of agency, claims about what good agents are like. Good agents are just those who perform the function of agency as well as it can be performed. Just as a good house is one that shelters well, and a good knife is one that cuts well, likewise a good agent is one that performs its constitutive activities well. We could hardly ask any more of agents, just as we could hardly ask any more of knives or houses. [7]

Furthermore, this strategy promises to be skeptic-proof: a skeptic who purports to doubt whether a knife ought to cut well is someone who just misunderstands what a knife is. Likewise, the thought is, a skeptic who purports to doubt whether agents as such ought to X well, where "X" specifies the function of agency, is someone who just misunderstands what agents are.[8]

My task in this chapter is twofold. The first task is to show why appeals to the idea that agency has a "function" in fact do not help the constitutivist's derivation of "ought" from "is." As I argue, appeals to "functions"—and relatedly, to the idea that agency has an end or "telos"—are compatible with

[7] Korsgaard 2009, ch.2.
[8] Korsgaard 2009: 29-30.

anti-constitutivism. The notions of "function" and "telos" are, in and of themselves, idle wheels in constitutivist arguments. Everything turns, instead, on how these notions are interpreted in the context of a theory of agency: it is the nitty-gritty of the theory of agency that matters for the success of the constitutivist derivation, not whether the notions of function and telos can be used to frame that theory. The point is not just the obvious one that the nature of agency matters for standards of practical reason if the constitutivist derivation of "ought" from "is" is any good. Rather, the point is that whether the derivation *is* any good depends upon the nature of agency. This is my argument in §1.2. The argument makes it urgent for constitutivists to articulate some other argument schema, besides the kind of "function" argument I repudiate, for exactly how the derivation of "ought" from "is" is supposed to go. (Alternatively, if we prefer to continue talking in terms of functions and excellent functioning, the invalidity of the function argument as it stands at least requires us to supplement it somehow.[9])

My second task is then to achieve this articulation of how constitutivist arguments at their strongest actually work. Constitutivists themselves have not, I think, articulated as strong an argument schema for the derivation of "ought" from "is" as can be done. This state of affairs can make it seem either that the derivation is subject to easy objections, or at any rate that constitutivism is not as powerful a theory as it is at its best. As I argue, however, given a certain kind of conception of agency—a certain kind of "is"-claim—the constitutivist's derivation of "ought" from "is" is unassailable. Tempting objections to the derivation cannot stand on their own, without seriously engaging philosophy of action. Not only this: at its strongest, the constitutivist argument threatens to *crowd out* all other options. It threatens to establish that if there are any constitutive norms of agency at all, then those are the *only* practical norms there could be: there is no room for hybrid views. The result is that anyone who cares about what the norms are by which we should guide our practical lives should also care about philosophy

[9] One supplementation would be a claim about the *kind* of function that agency has. But the argument would need to also make it clear why the standards of excellence for functioning of that kind derive from the nature of the function, not from something else. As we will see, these are essentially just the issues that a better constitutivist argument schema has to address, whether the argument is framed in terms of the notions of functions and excellent functioning or not. Thanks to Kieran Setiya for urging me to clarify this.

of action. Our only options are either to embrace constitutivism, in the strong form I articulate, or to defend a specific sort of account of agency on which the nature of agency fails to support constitutive norms. Either option requires defending theses in philosophy of action. So philosophy of action is inescapable for metaethics. This is the argument of §1.3.

My own position is ultimately anti-constitutivist. Indeed, I doubt that any practical norms can be given a skeptic-proof defense, whether by appeal to the nature of agency or by appeal to anything else. But I will not argue for that position in this chapter.[10] The point of the present chapter is just to show how constitutivist arguments at their best work, why no-one in practical philosophy can afford to ignore them, and so why no-one in practical philosophy can afford to ignore questions about the nature of agency.[11]

1.2 IF AGENCY HAS A FUNCTION, MUST PRACTICAL NORMS BE CONSTITUTIVE NORMS OF AGENCY?

1.2.1 "Function" and "good functioning"

The "function" argument outlined above (p. 8) purports to link the function or characteristic activity of agents to their *good* functioning, deriving standards for the latter from the former. In response, we might ask: even if this strategy is sound, what does it have to do with practical norms? Why think that goodness *qua* agent corresponds to good practical reasoning and reasons, so that the nature of agency, even if functionally specified, can yield an account of practical norms?

---

[10] The rest of my dissertation argues for it.

[11] Of course, if anti-constitutivism is right, then ultimately practical norms, if there are any, cannot be grounded in the nature of agency. And in that sense the nature of agency will be irrelevant to grounding practical norms. But the point is that, just like constitutivism, this anti-constitutivist conclusion is only warranted given justified commitments to controversial theses in the philosophy of action.

But the answer seems simple. In investigating agency, we are interested in specifically rational agents, in the minimal sense of 'rational' that contrasts with 'non-rational'. It is essential to agency in this sense that agency is the capacity to act for reasons and to deliberate towards action. This, at any rate, is the only kind of agency that practical norms could intelligibly govern. Practical norms just are norms that specify what good deliberation and good reasons for action are; hence, practical norms apply only to beings capable of deliberating and acting for reasons. Since, in evaluating the promise of constitutivist arguments, it is such beings we are interested in, it seems fair to say that deliberating and acting for reasons just is the function of agency, schematically speaking.[12] (We may take this specification of a schematic function as a definition of the sort of agency we are interested in.) If this is right, then it seems to follow that an agent who performs her function *well* deliberates well, and acts for reasons well, or as she should (provided her deliberations rest on no false beliefs).[13] An account of good functioning of agents then does seem to be an account of practical norms, in a quite trivial sense.

A more pressing objection is to the claim that, in identifying the function of agency, we already gain, at least implicitly, an account of what it is to perform that function *well*. It is obvious that to perform a function well must be, minimally, to perform it: so it must at least be relevant to an account of good deliberation that it is an account of, specifically, good *deliberation*. But this is just to identify the object of evaluation, the sort of thing to which the norms in question are to apply. Obviously the norms that apply to an object must in some sense "match" the object they govern: they must intelligibly be norms *for that object*. A peapod cannot intelligibly meet the norms for good deliberation, for it lacks the capacity for deliberation. So good and bad deliberators must share something: they must have the capacity for

---

[12] There may of course be cases of acting for no reason. The question whether such cases could be central to some interesting type of agency is not much discussed; but see Anscombe 1957: §20.

[13] It seems more natural to say that practical norms specify what good reasons for action are in given circumstances, than that they specify what good action-for-reasons is in given circumstances. And we might think that "acting for reasons as one should" is distinct from "acting for good reasons." Perhaps one may act for reasons well, or as one should (successfully?) even when the reasons for which one acts are bad by some further standard. I consider this kind of suggestion as an objection to constitutivism in §1.3. As I argue, given the right sort of account of agency, the gap between "successfully" and "for good reasons" necessarily closes; what the success of constitutivism then hinges upon is the correctness of some such account of agency. (Cf. Korsgaard 2008b: 210, fn. 5, where she clearly regards "being good at" acting for reasons as equivalent to tending to act for good reasons.)

deliberation, and they must exercise it. Moreover, in order to see what good and bad deliberation is, we must see how and why it is that certain kinds of deliberation are good, others bad. Plainly an account of practical norms must, then, attend to those features of good and bad deliberation that distinguish them from each other, as kinds of deliberation. But none of this entails that an account of the norms that specify when one's deliberation counts as good and when bad must be implicit in the nature of deliberation as such. The relevant norms may instead be imposed from the outside, where the question of their justification and content remains alive even after the question of what deliberation as such amounts to is settled.

That there is a potential gap to be bridged between an account of the "function" of agency and an account of the "good functioning" of agents is especially clear if by specifying the "function" of agency we mean merely to specify what is essential to agency, the activity or activities in which the active exercise of agential capacities consists. This "essential nature" reading of "function" seems to be the most innocuous one in the present context, since the constitutivist's ambition is to derive practical norms from the essential nature of agency.[14] Korsgaard herself seems to think of "function" in these terms: she says that when a thing performs its function, it does "whatever it does [and how it does it] that makes it the kind of thing that it is" (2009: 27).[15] But surely the fact that we can talk about the essential nature of agency in terms of the vocabulary of "functions" does not in and of itself secure the claim that practical norms are constitutive norms of agency. If it did, constitutivism would come cheap indeed. If there was a gap to be bridged between "is" and "ought" before, it is not closed by changing our terms.

Furthermore, whether there even *is* such a thing as functioning well or badly with respect to an object X is itself a fact that is external to the fact that the object has a function in the sense of an essential nature or activity, a something that it does and a way that it does it.[16] Some objects seem to have an

---

[14] It is not innocuous to someone who doubts that there are essential natures. But I assume that there are, and focus only on the constitutivist claim that practical norms are grounded in the essential nature of agency.
[15] Compare Korsgaard's discussion at 2008a: 134-140. I take "how it does it" from that discussion.
[16] I assume here that the essential nature of agency consists in an activity, and a corresponding capacity for that activity. In calling deliberation and acting for reasons 'activities', I mean to leave it open whether they also have an

essential nature or activity, and in this sense a "function," without even being intelligibly governed by norms. For instance, active main sequence stars essentially perform fusion; this distinguishes them and their characteristic causal powers from other heavenly objects, such as planets and neutron stars. What they do is fusion. When they stop doing that, they stop being active main sequence stars. But it is hard to see what it would mean to judge that something is bad *qua* a star, or that it performs fusion well or badly. It may do so at an especially high or low rate. But these are differences, not defects or excellences; and when the fusion reaction stops altogether, the star undergoes a transformation into something else. That is all. In Korsgaard's terms, the star's "form" changes when it stops performing its function: the star literally disintegrates, ceasing to be what it was (2009: 28). But whereas Korsgaard wants to say that literal disintegration is always at the extreme end of a scale along which there are, before it, increasing degrees of *badness* (ibid.), it does not seem apt to call stars bad *qua* stars. Talk of "functions" and "forms" does not seem to here support normative judgments at all.

All of this suggests that appeal to "functions," when a thing's "function" is understood merely as a specification of the essential nature or activity of a thing, might be an idle wheel in constitutivist arguments. In and of itself, it does not help us see how an account of "good functioning" might be derivable from a specification of the function of agency. However, a natural response on behalf of the constitutivist is to appeal to an accordingly stronger notion of "function": perhaps there is after all more to "function" talk than mere talk of essential natures.

In particular, the constitutivist might note that the idea that an object has a function is at home in a *teleological* conception of the object—in a conception of the object as essentially serving some *telos*, often translated either as *purpose*, *end*, or *good*. Narrow focus on what, say, a heart does on its own might lead us to identify its function as simply that of pumping blood. But we do not properly understand *what this is* without viewing it in the context of what it is *for*—in the context of the entire organism whose

---

"end" or *telos*, whether internal to or external to the activity itself. (Cf. Aristotle's idea that all (human) activities seek some "good" or "end," and his distinction between *productive* activities, whose "end" is external to the activity itself (such as bridle-making), and activities whose end is instead internal to the activity itself: NE I.1.) I discuss the idea of a *telos,* and its connection to the notion of function, below.

flourishing the heart subserves as an end. We understand what a heart is only when we understand how its proper functioning contributes to the whole organism's healthy, flourishing life. The heart performs its function by producing a cardiac output—and not just any cardiac output, but one that enables the whole organism to live. But by the same token, we get standards for what a good heart is: what a good cardiac output is for a given heart depends on what a healthy, flourishing life is for the type of organism in question.

Supposing there is something to this, the constitutivist might then try linking the idea of a thing's "function" to the idea that it has some *telos*, so that there is after all more to talk of "functions" than merely talk of essential natures. In particular, perhaps the notion of function is ill-applied precisely when the sort of object in question has no essential *telos* by reference to which we could make normative judgments about how good or bad an individual object is at achieving the *telos* for objects of its kind. On this suggestion, the reason why stars can have essential natures without thereby supporting normative judgments is just that their essences do not amount to "functions" in the requisite sense. The essential nature of a star is not of the right *sort* to count as a function, and so not of the right sort to support normative judgments. This is why the star lacks constitutive norms. Still, all things belonging to *functional* kinds do have their own constitutive norms, because to belong to a functional kind just is, *inter alia*, to have an essential *telos* or purpose or good the achievement of which sets standards for the good or bad performance of the thing's function.

This pairing of "function" talk with the idea of a *telos* is indeed a natural strategy for constitutivists to try. But of course, the idea that agency has a *telos*, some good at which it essentially aims, is already a much more tendentious thesis about the nature of agency than is the bare idea that the essential nature of agency is, schematically speaking, a matter of having and exercising capacities for deliberation and acting for reasons. The tendentious issues here are: What is the "good" or "end" in question that agency as such essentially aims at? And what is it to "aim" at it? And why think that agency

as such has, or aims at, any "good" or "end" at all?[17] In the case of an organ such as a heart, we get a *telos* into view in part because the organ is merely a part of a larger organism. The organ literally does not function at all outside the context of the organism. In that sense, the organ as such is something that subserves the whole organism's functioning; and the organ's good functioning subserves the flourishing or healthy life of the organism. But of what organism or larger thing might rational agents as such be parts? It may be that rational agents cannot *arise* except in the context of a society of rational agents. But this genealogical connection does not yet entail that rational agents cannot function at all outside the context of some such society, nor that they "subserve" the society in any relevant sense yielding a notion of "good functioning." Compare stars: they too have a genealogical connection to their environments: they need a star-forming region, a dense molecular cloud, to be born. They do not come into being on their own. But they can function as stars once outside such regions. And as we saw, it seems odd to say that stars have a *telos* in the sought-after sense that would support normative judgments.

Similar issues arise with regard to purported artifactual analogies. Tools are often mentioned as prime exemplars of objects that have a *telos* in the sought-after sense. But that is because they are designed (or adopted) for a purpose: the purpose or *telos* is, roughly, decreed by the user or users of the tool. That is why the tool is in part constituted as the tool it is by the purpose decreed. But an analogous thesis about the *telos* of agency would be hugely controversial.

To be sure, there are interesting issues to explore here. But the point is that defending or repudiating the thesis that agency has a *telos* in a sense that supports normative judgments already belongs to the nitty-gritty of the philosophy of action: what the thesis might mean and why we should accept it are questions to be answered in the context of a detailed defense of a theory of agency. The thesis is not achieved by any obvious schematic characterization of agency that all constitutivists and non-

[17] For constitutivism framed in terms of constitutive "aims," see Velleman 2000, 2009; though Velleman denies that the constitutive "aim" of action is properly thought of in terms of a positive evaluation, or as a "good". The constitutive "aim" of action is instead "what makes sense" (2000: 121). I discuss below the precise sense in which the idea that agency as such "aims" at some *telos* might help the constitutivist.

constitutivists alike accept. If the notion of a *telos* validates the derivation of "ought" from "is," then the derivation is already dependent on in-depth theses in the philosophy of action.

Furthermore, as I will argue next, even if agency as such does aim at some *telos*, this idea can be given a non-constitutivist interpretation in the context of a theory of agency—though as we will see, this depends on what the *telos* is and what it is to "aim" at it. This will complete my argument that the notions of "function" and "telos" do not, in and of themselves, help the constitutivist derivation of "ought" from "is" at all. The constitutivist derivation is more clearly and effectively formulated and defended directly in terms of the specifics of the sort of conception of agency from which practical norms can putatively be derived.

1.2.2 "Having" and "aiming" at a *telos*

My argument here is in two parts. First, I identify a criterion for the sense in which agency must have a *telos*, if the *telos* is to be a source of norms for good functioning. Second, I argue that even if agency does have a *telos* in that sense, there are non-constitutivist interpretations of what it is for agents as such to "aim" at it. This yields a second criterion for how constitutivists must interpret "telos"-talk for it to serve their purposes.

A comment on Korsgaard's take on "telos"-talk yields the first criterion. Despite her aforementioned identification of a "function" with a thing's essential nature, Korsgaard does elsewhere suggest that having a function in a sense that supports the idea of *mal*functions—and so in a sense that supports normative judgments concerning good and bad functioning—is connected to a thing's having a "purpose" or *telos* at which it constitutively aims (2008a: 141; cf. 2009: 35). But Korsgaard's ultimate *attitude* to "telos"-talk is odd. In explaining and justifying her "teleological thinking"—the idea that each object has some *telos*, and a function in the associated sense—Korsgaard emphasizes that her teleological claims are only supposed to individuate objects as (kinds of) causally interesting regions of the manifold for us *qua* cognizers of those objects. We need to conceive of each object as serving some *telos*,

Korsgaard thinks, to cognize it as an object at all. But, she says, "there is no claim here that everything has one and only one purpose that is in fact its natural purpose" (2009: 37-9). Her aim in making this concession is to ease worries that invoking teleology conflicts with the "Modern Scientific World View" (37). The concession casts doubt, however, on how seriously a metaphysics of agency should take the identification of particular "purposes" rather than others. And this in turn makes trouble for the constitutivist strategy. If a thing is to be judged good in relation to its purpose, it is to be judged good in relation to *what the purpose actually is*. Unless a purpose is some actual purpose, we cannot hope to derive actually authoritative practical norms from it. But unless we have reason to take specific proposals about the "purpose" of agency seriously, we have no reason to take any norms that may putatively derive from that purpose seriously, either.[18]

Here, then, is the first criterion for the sense in which agency must have a *telos*, if the *telos* is to be a source of norms for good functioning. The *telos* proposed must be some substantive *telos*, and it must be non-optional. There are two thoughts here. First, what good or bad functioning amounts to for a thing depends on the substance of what the thing's *telos* is, and (relatedly) what it is for that thing to serve or aim at it. The mere notion of serving or aiming at *some* telos *or other* does not yet yield actual, contentful standards for good or bad functioning. Second, that the *telos* must be non-optional means that we must have reason to take that particular proposed *telos*, and the standards that putatively derive from it, seriously. (Constitutivists who wish to rely on the notion of a *telos* must think that the reason to take a particular proposed *telos* seriously as a putative source of standards for good functioning is that aiming at the *telos* in question is somehow a necessary feature of the metaphysics of agency.)

However, even if agency does have a *telos* in that sense, there are non-constitutivist interpretations of what it is for agents as such to "aim" at it. Seeing why will yield the promised second criterion for how constitutivists must interpret "telos"-talk for it to serve their purposes.

---

[18] Furthermore, the example of stars, raised above, seems to be a counterexample to Korsgaard's claim that cognition of objects *is* necessarily of objects as serving a *telos* that supports constitutive norms.

We can articulate a non-constitutivist interpretation in the context of Aristotle's *Nicomachean Ethics* (NE), with reference to how the idea that *eudaimonia* is the *telos* of human action figures in that work. It is clear that Aristotle has something quite substantive in mind with *eudaimonia*, the human *telos*, that he wishes to elucidate, and in relation to which all actions and undertakings are to be judged good or bad. It is also clear that he thinks it will be helpful, in elucidating this, to understand the human function; for a thing's good in general consists in its performing its function well (NE I 1097b23-1098a18). In particular, according to Aristotle, (1) the human function is rational activity (NE I 1098ab33-a14); (2) performing this function well is performing it in accordance with the excellences related to it, since "excellence" in general is the technical notion of a characteristic or disposition or state that makes a thing perform its function well (NE I 1098a11-18; II 1106a15-24); and (3) when the human performs its function well, it achieves its *telos*, its good, which everyone agrees is *eudaimonia* (NE I 1098a8-18). (4) What Aristotle then does throughout Books II-VI is to make some rather specific claims about what the excellences are; and he evidently takes himself to thereby elucidate the substance of *eudaimonia*—of what rational activity in accord with the excellences comes to.[19]

Now, nothing in these four points yet commits us to the constitutivist claim that an account of performing the human function well is implicit in an account of what it is to perform it, as such. All that follows thus far is that, in order to see what it is to perform the human function well, we must see what the function is whose good and bad performance is in question, and how its good performance (and so its excellences) differs from its bad performance (from the defects and excesses). This is compatible with thinking that an account of the excellences, and of how they make the human perform its function well, is not itself implicit in our account of the function. So Aristotle's four points are thus far compatible with thinking that an account of the actual substance of the *eudaimon* life—of the *telos* of human action—is

[19] I assume the plausible view that, for Aristotle, rational activity in accord with the excellences just amounts to *eudaimonia*, instead of the view that *eudaimonia* is some further product of that activity, not yet characterized in substance by characterizing what activity in accord with the excellences is. Various things that Aristotle says in NE I support this plausible view: most blatantly, the conclusion of his "function" argument, namely, that "the human good turns out to *be* activity of soul in accordance with excellence [etc.]" (1098a16-17; my italics).

not derivable from, or an essential part of, an account of the human function as such. Since it is the actual substance of the *eudaimon* life that sets the standards for good functioning, it follows that the standards for good functioning need not, as yet, be implicit in an account of what it is to perform the function, just as such.

However, Aristotle does also assert a fifth thesis, namely that (5) not only good human actions but *all* human actions and undertakings in some sense aim at the human *telos* (NE I 1094a1-22). This suggests that the human function is itself essentially related to, or partly defined in terms of, the human *telos* in some way (since all actions and undertakings aim at it): the human *telos* is, we might say, the "constitutive aim" of all of our actions, so that one cannot perform the relevant function *at all* without aiming at the *telos*. Does this imply that an account of the substance of the *telos* itself, and so of the standards in relation to which exercises of agency are good or bad, is internal to the function of agency?[20]

No, it does not. Aristotle's fifth thesis can be interpreted non-constitutivistically.[21] I will first present an argument to this effect, and then qualify it in important ways, yielding our promised second criterion for how "telos"-talk must be interpreted by constitutivists.

Suppose, for now, that "aiming" at something is a matter of desiring or intending it, or of having it as your goal.[22] Now, in this sense of "aiming," action as such might "aim" at *eudaimonia* merely in the sense that all action must be ultimately motivated by a desire or inclination for *eudaimonia* under *some conception of eudaimonia or other*. One may aim at *eudaimonia*, in the sense of aiming at "living well and doing well," even if one's specific conception of what this would come to is false (NE I 1095a18-25). Indeed, something like this may be what is distinctive about the bad person: although she too acts for the sake of *eudaimonia* or the chief good, she goes wrong in pursuing it because her specific views about what things are worth doing or pursuing, under the heading of *eudaimonia* or the chief good, are

---

[20] Since we are discussing constitutivism, the thesis that the nature of rational agency as such grounds practical norms, I ignore Aristotle's focus on humans in particular rather than rational agents more generally.

[21] I will not argue that this is how we should interpret Aristotle. (Though see fn. 23 below for some moves in this direction.) My interest is only in the claim that there is a possible interpretation of "aiming" at the *telos* of agency that is not constitutivist.

[22] Other possible senses of "aiming" arise at the end of §1.2.2.

mistaken. (This in turn is in large part because of bad upbringing, which has instilled bad character traits—defects or dispositions to excess instead of excellences—traits that make the bad person's views about what is worth pursuing false; NE II, VI 1140a24-b30.) Since the bad perform their function badly, and so reason and act badly, the sense in which *they* aim at *eudaimonia*—namely, under a false conception of it—cannot correspond to the standards for being a good agent. It is only the correct conception of *eudaimonia*—the actual substance of *eudaimonia*, as it really is—that corresponds to good reasoning and action. Hence an account of the standards of practical reason is not retrievable from the claim that everyone must "aim" at *eudaimonia* under some conception or other. What the correct conception of *eudaimonia* is, and so what the standards of practical reason are, is just a matter of how *good* agents act. But this we already knew. It does not help us get closer to an actual account of those standards.[23]

It is an important question in what sense one still aims at *eudaimonia*, and not something else, if one's conception of it is *completely* wrong. I come back to this question below. For now, let me consider an objection to the argument just presented. It will yield a caveat to the argument, and help us formulate the promised second criterion for how constitutivists must take "telos"-talk for it to serve their purposes.

---

[23] Korsgaard gives a contrasting constitutivist reading of Aristotle in her 1986a; cf. 2008a. Korsgaard seems to think there could be a non-constitutivist interpretation, but seeks a constitutivist interpretation on the ground that Aristotle's argument would be more powerful interpreted constitutivistically: a constitutivist interpretation promises to explain why Aristotle's various normative claims about the substance of *eudaimonia* are justified, where their justification would otherwise be left obscure. As an interpretive matter, I think there are in fact some features of NE that strongly suggest a non-constitutivist interpretation. First, Aristotle never claims to derive a specification of the excellences from a specification of the function of (human) agency—something one would expect him to advertise prominently, were he engaged in attempting such an exciting and ambitious feat. (Cf. McDowell 1995b: 207-8.) Second, Aristotle explicitly addresses NE to an audience that is well brought-up—that is, to an audience who already shares a grasp of "the that," the specific things that are correct in the light of *eudaimonia*; he evidently thinks that it would be pointless to address a badly brought-up audience (i.e. people who would be liable to be skeptical about the specific first-order normative claims Aristotle makes) (NE I 1094 b28 - 1095 a13; Cf. McDowell 1995b: 212). This is significant, because even if there are no guarantees that any particular skeptic would be satisfied by even a sound constitutivist argument, it would be clearly unwarranted to think that people whose moral upbringing is bad *must* also be thick when it comes to theoretical argumentation about the metaphysics of (human) agency, and about what normative claims the nature of agency entails. So were Aristotle attempting a constitutivist argument, the restrictions he places on his audience would be strange and unwarranted. I think Aristotle's strictures instead point to the view that he regards the substance of *eudaimonia* as an object of irreducibly normative investigation, not as something to be derived from an account of (human) agency as such. But I will not press these points here.

Here is the objection.[24] Suppose that agents as such must "aim" at knowledge. We need not be able to derive an account of what knowledge is from this thought. Still, given that there is such a thing as knowledge, the fact that agents as such must aim at it sets an objective standard for their actions: actions go well if they maximize, or otherwise achieve, knowledge. This is still a constitutivist view. The fact that all actions must aim at knowledge anchors the standard of maximizing knowledge—whatever maximizing knowledge amounts to—to the nature of agency, giving us an explanation of why we should care about maximizing knowledge. A skeptic might have previously doubted that knowledge-maximizing has anything to do with good practical reasoning and reasons, but no longer. If she understands what agents are, she understands that they necessarily seek knowledge; and since good agents must then do so *well*, the skeptic cannot consistently doubt that agents ought to hit the target they necessarily aim at. (This leaves it open that there may be *more* to "well" than actually hitting the target.[25] But it is at least implausible to hold that one could seek knowledge "well" while entirely missing this target.)

In response, I think we should concede that a non-constitutivist interpretation of "telos"-talk of the sort I proposed above is not possible for all proposed *telē.* This is the caveat, in outline. So what must be the case for such a non-constitutivist interpretation to be possible? And what does this tell us about what constitutivist interpretations must be like?

Consider what Aristotle says the indisputable content of *eudaimonia* is: "living well and doing well" (NE I 1095a20-21). He then interprets this, *via* identifying the function of humans as rational activity, as rational activity in accord with the excellences (NE I 1097b25-1098a23). *Eudaimonia* just is, or at least includes, doing well *qua* rational agent.[26] We then face the question of what doing well *qua* rational agent actually amounts to—what the standards of practical reason are. (This is the question we have been calling the question about what the actual substance of *eudaimonia* is.) Now, clearly it would

---

[24] Kieran Setiya put this objection to me; I hope I have not distorted it too badly.
[25] §1.3.2 discusses the possibility of "hybrid" views, on which there are both constitutive and non-constitutive norms.
[26] I ignore possible issues concerning the relation between the contemplative and practical lives in Aristotle's NE, assuming that we can treat of the practical aspect of *eudaimonia* on its own.

not be informative in the least to say that, to meet the standards of practical reason, we ought to do well *qua* rational agents (that is, be *eudaimon*). This would just be to repeat the trivial truth we started with at the outset of §1.2: that doing well *qua* one who deliberates and acts for reasons is to meet the standards of practical reason. So the constitutivist can hardly claim to have explained why *eudaimonia* is relevant to the standards of practical reason by noting that all agents as such aim at it. Given Aristotle's interpretation of *eudaimonia* as excellent rational activity, there is simply no question why we should take this proposed *telos* seriously as a source of standards for reasoning and acting well. If there are standards of practical reason, *eudaimonia* includes them. That is why knowing how to be *eudaimon* would be of obvious relevance for knowing what the standards of practical reason are. Unlike in the case of knowledge, the additional claim that agents as such aim at *eudaimonia* in all of their actions is irrelevant to explaining the normative relevance of *eudaimonia*.

Correspondingly, while the claim "Agents as such must seek knowledge well to be good *qua* agents" is non-trivial (given that "knowledge" is not interpreted just as "deliberating and acting well"), in contrast "Agents as such must do well and live well in order to be good *qua* agents" *is* trivial, given that being good *qua* agent is just an aspect of doing and living well for creatures like us.[27] Anchoring *eudaimonia* to the nature of agency by saying that all agents as such aim at it does absolutely *no* work in getting us towards an account of what the standards of practical reason actually are. It is still a completely open question what reasoning and acting well comes to.

At least, this is so *unless* at least some of the specifics of the correct conception of *eudaimonia* are retrievable from the mere concept, or from the mere fact that agents must "aim" at it under some

---

[27] This is of course compatible with the thought that it is *not* trivial to claim that reasoning and acting well—excellent rational activity—*are* aspects, or even the most central aspects, of doing well and living well for creatures such as us. This is why Aristotle gives his "function argument": it yields an outline account of doing well and living well for creatures like us. But the fact that "reasoning and acting well" constitutes an outline account of "doing well and living well for creatures like us" does not entail that "doing well and acting well for creatures like us" constitutes an outline account of "reasoning and acting well," and so of the standards of practical reason—at least when "creatures like us" are just agents.

conception or other. Can we make sense of the idea that agents as such might aim at being good *qua* agents, but under a *completely* false conception of what this amounts to?

I think we can. We must of course assume that the false conception is still at least a conception *of* doing well *qua* agent. But with this proviso, it certainly *seems* possible to have a completely false conception. Suppose we interpret "aiming to be good" as saying nothing more than that, in deliberating and acting for reasons, agents must always take the reasons for which they act (or their "premises" in deliberation) as counting in favor of their actions in some way or other, even if their conception of which considerations in fact count in favor of which actions is completely false. This is a (very quick, but for our purposes sufficient) version of the claim that agents must act "under the guise of the good." It is not a trivial claim about agency: some have denied that any version of it holds.[28] Nonetheless, it provides an interpretation of "aiming" at *eudaimonia*; and one that does not, on its own, seem to help us get any closer to an account of what it is to actually *be* good *qua* agent. For it seems, prima facie, that different agents might take, say, the fact that φ-ing would be cruel as either a consideration *in favor* of φ-ing or a consideration *against* φ-ing. Such agents certainly have very different conceptions of what counts in favor of what. Assuming both conceptions cannot be true at once, a completely false conception of at least a *region* of practical standards seems possible. If one's conception of *eudaimonia* could not be likewise faulty in *any* region, or indeed in all regions at once, it is not immediately evident why.

The issues raised here, about the possibilities of agency, are of course complex. The purpose of these brief remarks is merely to gesture at the possibility that "aiming" at *eudaimonia*, when such "aiming" is understood as a thesis about the guise of the good, might leave it entirely unsettled what the standards of practical reason are. And if "aiming" at *eudaimonia* in the "guise of the good" sense does *not* leave it unsettled what the standards of practical reason are, then this is because of the intricacies of the theory of agency, not because of the very notion of "aiming" at *eudaimonia* in the sense introduced: it is

[28] Setiya 2007, Part I; Setiya 2010.

an issue about whether completely and globally conflicting ways of taking one's reasons to be good, when one acts on them, are possible.

One might object that even the very notion of "aiming" at *eudaimonia* in the introduced "guise of the good" sense does inevitably import at least some standard of practical reason, however thin. For if agents must always take their reasons for action to be in some way good, then it is at least a standard of practical reason that one must take one's reasons to be in some way good. In response, however, if *any* action for a reason is action for some reason one takes to be in some way good, then one could not *fail* to act for some reason one takes to be in some way good, while still acting for a reason. And that means that acting for reasons one takes to be in some way good cannot be a genuine norm governing action for reasons: where there is no possibility of discord with the putative norm while still acting, there is no normativity either.[29] Taking a reason to be in some way good is just a description of what happens whenever one acts for a reason. One either does it, and so acts for a reason one takes to be in some way good, or one does not, and so does not act for a reason at all. The guise of the good, on the above interpretation, does not yield any practical norm, however thin.[30]

Of course, the proposed interpretation of "aiming" at *eudaimonia* might itself be too thin to be plausibly attributed to Aristotle. Moreover, it might be odd to continue thinking that "aiming" at *eudaimonia* is having it as an object of *desire* under some conception or other. But what matters for the present argument is just that there is some available conception or other of "aiming" at *eudaimonia* that does not, in and of itself, yield any account of the standards of practical reason, even in outline. Perhaps *eudaimonia* is instead the object of *intention*, in that whenever one acts for a reason intentionally, one intends to be acting for that reason, conceived of as in some way a good one. Further, we might suggest

---

[29] This "violability" requirement on normativity is noted by e.g. Korsgaard 1997: 228 and Dreier 1997: 91.

[30] A different interpretation of the guise of the good might hold that one must always be at least disposed to act on one's best judgments concerning what one ought to do. We might then further hold that good agents perfectly or successfully manifest this disposition, whereas bad ones do so imperfectly or badly. On this view, *enkrateia*, acting as one judges one ought, would be a standard of practical reason. But I think we can reject this interpretation of the guise of the good. Often enough, practical reasoning seems to involve no such judgments. I say much more about practical reasoning in chapters 2 and 3.

that which reasons an agent takes to support which actions in turn evinces, at least in part, that agent's conception of *eudaimonia*—however faulty that conception might be. We need not go into the details of such proposals to see that a non-constitutivist could make them. Every action done for a reason can aim at the *telos* of agency, at doing well *qua* agent, even if one aims at it under a completely false conception of what doing well *qua* agent actually is.

If the foregoing is right, then to make use of "telos"-talk to elucidate the derivation of "ought" from "is," constitutivists must hold that even the most summary specification of the *telos* of agency is not just some version of "doing well *qua* agent." Otherwise it looks to be possible to aim at it under a completely false conception—so false that the fact that agents necessarily aim at it under some conception or other is rendered normatively empty. This is the second criterion for how constitutivists must interpret "telos"-talk for it to serve their purposes. It is not enough that the *telos* is something substantive, of non-optional import for specifying standards of being good *qua* agent, and something that agents as such necessarily aim at. It must, further, be impossible for agents to "aim" at it under a conception so false as to empty the fact of everyone's necessarily "aiming" at it of all normative import.

In §1.2, I have been examining the type of "function" argument constitutivists such as Korsgaard (2009) and Smith (2010) employ to explain the derivation of "ought" from "is." The examination led us to identify a gap between talk of functions and talk of good functioning. Talk of *telē* was supposed to fill that gap. Our conclusion is that it does not do so automatically. Whether it does depends upon what the purported *telos* is and what it is for agents as such to "aim" at it. Even if agency has a function and a *telos*, this is insufficient on its own to validate the constitutivist derivation of "ought" from "is." All the work is done by the details of the account of agency that the notions of "function" and of "aiming" at a "telos" are used to frame.

This makes it urgent for constitutivists to articulate some other argument schema, besides the "function" argument, for exactly how the derivation of "ought" from "is" is supposed to go.[31] My task in the next section (§1.3) is to formulate such an alternative schema for the constitutivist derivation. I show that the schema is generally valid, given an outline conception of agency that all constitutivists can share. And I argue that this has some surprising consequences.

1.3 THE TRUE STRUCTURE AND APPEAL OF CONSTITUTIVIST ARGUMENTS

1.3.1 Spelling out the argument

Besides "function" arguments, there is a different strand of argumentation in Korsgaard's (2009) and Smith's (2010) work, a strand that they connect to the function argument, but that can be given a stronger and clearer articulation on its own. This strand makes use of the idea that there is a specific principle or principles of reasoning such that following the relevant principle(s) is metaphysically necessary and sufficient for practical reasoning as such. Correspondingly, one acts for reasons if and only if the transition *from* one's entertaining a putative reason or reasons *to* one's action is guided by the principle(s) in question. The idea is then to try to argue from this fact about agency to the normative authority of the relevant principle(s). To reason in accord with the relevant principle(s) perfectly or well just is to meet the practical norms that authoritatively govern agency.

As I hope to show, an argument of this type promises to be particularly powerful. If some instance of the type of conception of agency on which the argument is premised is true, then the derivation of "ought" from "is" is unassailable: tempting objections to the derivation cannot succeed on

[31] Though alternatively, one might supplement the "function" argument. (See p.9, fn.9.) Certainly the schema I will propose could be construed as such a supplementation: one might continue to frame questions about the nature of agency in terms of the vocabulary of functions, and questions about good deliberation in terms of the vocabulary of good functioning. But I think this would be superfluous.

their own, without objections to the premise about agency. Not only this: given the premise about agency, the argument leaves no room for any extra, non-constitutive norms. The result is that anyone who cares about what the norms are by which we should guide our practical lives should also care about philosophy of action. Our only options are to either embrace constitutivism, in the strong form I articulate, or to defend an account of agency on which its crucial premise is false. Either option requires defending theses in philosophy of action. Philosophy of action is inescapable for metaethics.

To explain the argument, let me first set up the terms of the discussion more thoroughly in terms of two basic steps in the constitutivist's argument. I will then lay out the argument schema for the constitutivist's derivation. §1.3.2. addresses an important objection. This will help to clarify and strengthen the schema.

The constitutivist's basic idea is to first argue that, to deliberate and act for reasons at all, one must follow a specific putative practical norm R. This is a specific thesis in the philosophy of action. For Smith, the relevant norm is an instrumental norm. As he puts it, the "function" of the psychology of agents is to produce action, and (according to his theory of action) producing action is a matter of combining "intrinsic"—that is, non-instrumental—desires with means-ends beliefs in accord with the following instrumental rule ME:

> (ME) Reason requires that (If a subject has an intrinsic desire that p and a belief that he can bring about p by bringing about q, then he has an instrumental desire that he brings about q [and this instrumental desire disposes him to actually bring about q].) (2010: 124)

We need not worry too much here about the proper formulation of this rule.[32] The important thought is just that there is a specific way in which agents must make transitions from their psychological states to action—in short, a specific way in which they must reason practically or make considerations their

[32] One might object that rules one "follows" ought to be formulated imperatively, not in terms of claims about rational requirements, or indeed, claims about anything at all. But for every rule imperatively formulated, we can substitute a corresponding principle that claims that the rule is a rational requirement, is reason-giving, or is in some other way authoritative. (For an example of such a procedure in the case of epistemic rules, see Boghossian 2001: 235-236.) To ease exposition, I will continue to talk about rules as "claiming" things, and I will use 'principle' and 'rule' of reasoning interchangeably. I will also say, interchangeably, that a rule's claims can be authoritative, or that they can be true. Nothing in the argument that follows will turn on these choices.

reasons for action—if they are to reason practically and act for reasons at all; and this specific way is roughly captured by the idea that agents must follow ME.[33]

Smith also tentatively suggests that there might be specific intrinsic desires that are part of agents' psychologies as such (2010: 133ff). If this suggestion were right, then one's psychology's being governed by ME would not be sufficient for agency. One could not yet act for reasons just by "following" ME.[34] However, if there are no necessary intrinsic desires, then I take it that Smith's thought is that it is not only necessary but also sufficient for agency that agents follow ME in their reasoning. Of course, there must always be *some* psychic raw material for the principle to engage—some desires and means-ends-beliefs or other. This is an enabling condition for our following ME at all. But so long as there is some such psychic raw material, following the principle is necessary and sufficient for action. It, and nothing more, is required.[35]

The second stage in the constitutivist argument is then to argue that the principles, whatever they are, that describe agents' psychologies as such, are thereby also authoritative requirements of reason. It is not just that agents must, metaphysically speaking, follow (say) ME in their reasoning; but further, ME is *true*. ME describes not only a metaphysical "must," about exercising practical reason, but also a normative "must," about the standards of practical reason.

Smith's ME is just an example. The same basic two-step structure of constitutivist argumentation is present in Korsgaard's work. Korsgaard (2009) thinks that, to deliberate and act for reasons at all, one must deliberate in accord with Kantian categorical and hypothetical imperatives. (Again, the formulations of these principles will not concern us here.) Each principle corresponds to a necessary, jointly sufficient,

---

[33] Dreier defends essentially the same idea in his 1997. For the idea of reasoning as practical when it involves transitioning from thought to action, see Anscombe 1957: §33.

[34] In such a case, we might wonder what "following" ME comes to: since it cannot be a case of practical reasoning, in the sense of reasoning that issues in action, what is it? This issue will come to the fore at the end of §1.3.2.2. We can put it aside for now.

[35] Why is it not a further metaphysical requirement of agency—also to be transformed into a normative requirement—to *have at least some desires and means-ends beliefs or other*? Smith does not consider this question, but I take it that his answer would be that, if one is an agent at all, this purported requirement is impossible to violate. And again, violability is a condition of normativity.

but individually insufficient, stage in deliberation: one must go through both stages to fully deliberate towards action. According to Korsgaard's theory of agency, the first stage is the formulation of a maxim of action: a proposed act-for-the-sake-of-an-end. To formulate a maxim, we must follow the hypothetical imperative. The second stage is then to test the maxim for whether its form is universal in a specific sense. This stage consists of following the categorical imperative. In going through both stages, we reason towards action. If one stage is missing, then practical reasoning as such is incomplete, and action cannot ensue. If both stages are present, then action ensues. That is, in a nutshell, Korsgaard's account of agency.[36] The next, second step in her constitutivist argument is by now familiar: to argue that the principles that describe deliberation and acting for reasons as such thereby also have normative authority.

The issue of how the constitutivist derivation can be validated is just the issue of how one can derive the constitutivist's purported normative "must" from her purported metaphysical "must." The argument schema I will outline explains this derivation. It will be crucial to this argument that following the principle(s), whatever they are, that purportedly describe the metaphysics of agency, is both necessary and sufficient for deliberating and acting for reasons at all. Without that crucial claim, constitutivist derivations face severe problems: with it, however, they are bound to succeed. At least, they are bound to succeed given the further condition that the principles of reasoning in question are, even though metaphysically necessary, also *violable*. Violability is a condition of normativity: if one simply cannot go wrong in one's reasoning by a principle's lights, then it is hard to see in what sense we could say that the principle really requires anything of one.[37]

One final preliminary: how can a principle be both metaphysically necessary for agents as such, and yet violable, so that agents might act and reason in accord with it *without* conforming to its dictates? The idea of a *disposition* to follow the principle helps here. Consider the example of following the rules of English grammar and the rules constitutive of the meanings of words in English. Although I am in general disposed to follow these rules, I nonetheless sometimes make performance errors—as, for

[36] Korsgaard 2009: 59-72, 81-90.
[37] Cf. e.g. Korsgaard 1997: 228 and Dreier 1997: 91.

instance, when I am really tired. But insofar as I speak English, I am still following the constitutive rules, however imperfectly: my disposition is manifesting itself, although less than fully or ideally. Something (such as tiredness) impedes. Likewise, the thought is, if I am disposed to reason in accord with R, I am an agent. This disposition must manifest itself to at least some extent whenever I reason or act for reasons. But it may manifest itself less than fully: in these cases, I go wrong by R's lights, just as I go wrong by the lights of the rules of English when I make performance errors. Still, so long as my errors are not *too* egregious, I am still intelligible as following the relevant rule.

Note that the idea of following a rule is flexible enough to accommodate those constitutivists who wish to think of agency as having a *telos*. For instance, if the *telos* of agency is knowledge, and good agents maximize knowledge, then agents as such can be represented as following the rule "Maximize knowledge!" Good agents then follow it well, bad agents badly. In more intuitive terms, we may put such a view by saying that the good inferential transitions from considerations *p* to actions φ are, roughly, the ones that (tend to) maximize knowledge (in the circumstances). In being necessarily disposed towards maximizing knowledge, then, agents may be said to "aim" at the *telos* by *approximating* to it. The framework of dispositions of reasoning, outlined above, therefore looks to be one within which all constitutivists alike can frame their views.[38]

Now for the derivation of "ought" from "is." In the argument schema below, 'R' refers to whatever putative norm or norms the constitutivist claims to be constitutive of deliberation and of acting for reasons, in the sense that following R is both necessary and sufficient for these activities. I will shorten 'deliberation and acting for reasons' as 'deliberation'. In discussing the argument in what follows, I will call it 'The Derivation'.

1. To deliberate just is to deliberate in accord with R, and in this sense to "follow" R. (The constitutivist's metaphysical "must.")
2. To deliberate well just is to meet all the practical norms that authoritatively govern agency. (Just as e.g. playing chess well is meeting all the norms for chess-playing.)

[38] Although I think this is basically right, some complications about "telos"-talk resurface below, in §1.3.2.2.

3. Necessarily co-referring terms are intersubstitutable *salva veritate* (in non-oblique contexts).[39]
4. If [1], then 'deliberation' and 'following R' are necessarily co-referring terms.
5. So, 'deliberation' and 'following R' are necessarily co-referring terms. (By [1] and [4])
6. So, to follow R well just is to deliberate well, and so it is to meet all the practical norms that authoritatively govern rational agency. (By [2], [5] & [3])

In other words, the conclusion is that R is a practical norm, and in fact *exhausts* the content of practical norms. There could be no further norms that agents are required to meet, since following R well *just is* to meet all the practical norms governing rational agency. (The constitutivist's normative "must.")

The Derivation looks pretty powerful, supposing that some instance of premise [1], the constitutivist's metaphysical "must," holds. Premise [3] seems true, [4] seems indisputable, and so does [5] if indeed [1] is true. The conclusion [6] follows from these, *if* [2] is true as well. We might capture the basic idea of The Derivation by saying that it transforms the "rule-internal normativity" of R—the fact that there is such a thing as following R correctly or incorrectly—to normative *authority*, *via* the fact that following R is metaphysically constitutive of deliberating and acting for reasons as such. In doing so, the argument answers the intuitively compelling question: "Any old rule can make claims about how we should deliberate and act, but which rule's claims should we really listen to and accept?"

If The Derivation works, it is quite exciting. For it claims not only that we can ground *some* practical norms in the metaphysics of agency, and so on a normatively non-question-begging foundation. That would be an exciting result in and of itself.[40] But The Derivation claims, more strongly, that we can in fact ground *all* the practical norms there are in the metaphysics of agency. If some instance of The Derivation is sound, construed so strongly, then there could be no gainful work left to do in defending

---

[39] By this I mean that, if term1 refers to A and term2 refers to B, and □(A=B), then term1 and term2 can be substituted for each other in statements without changing the truth value of the statement, so long as the statement is not the content of a propositional attitude. I put the point in terms of "necessarily co-referring terms," rather than "co-referring terms," to forestall the objection that contingently identical things (such as a statue and the lump of clay that composes it at time *t*) may not make true the same modal judgments. (I thank Kieran Setiya for this objection.) The constitutivist's metaphysical claim about agency is a claim about necessary, not contingent identity.

[40] I raise doubts later about the possibility of "hybrid" views, views admitting some constitutive norms and some non-constitutive ones. Jollimore 2005 suggests the possibility of hybrid views.

practical norms, after having gotten clear on the nature of agency and the constitutive norms it grounds. Constitutivism would be, so to speak, the only game in town.

The controversial premises are [1] and [2]. In the next section, I will address a tempting objection to [2]—or more properly, to the way in which [2] interacts with [1]. I argue that the objection fails given [1] itself. Furthermore, there are no other relevant objections to The Derivation that are not also objections to [1] itself. The consequence is that The Derivation inevitably goes through, given the specific sort of account of agency that [1] encapsulates. This yields the result advertised at the outset: our only options are either to embrace constitutivism, or to give an account of agency on which there is no "R" that makes [1] true. In either case, anyone who cares about what the standards of practical reason are cannot afford to ignore some rather in-depth questions in the philosophy of action.

1.3.2 Objection to The Derivation: equivocation on 'well' and the possibility of hybrid views

The objection is that the way that premise [2] interacts with [1] in The Derivation trades on equivocating on two senses of 'well'. We can motivate the accusation of equivocation in terms of the analogy with chess mentioned in premise [2].[41] In one sense of 'well', it is of course true that, in playing chess well, one meets all the norms for chess-playing. This is the sense of 'well' in which one plays chess well when one plays it *as well as it is possible to play*. In such cases, there is no room for claiming, truly, that one ought to have done something other than what one did.

But there is an ambiguity in 'well', corresponding to a distinction between two types of rule for chess. Some of the rules of chess seem to be constitutive rules: you cannot play chess at all without following them, and if you do follow them, you are playing chess. These are rules definitive of the legal moves within the game—rules such as "The bishop moves diagonally." If you tried to move the bishop

[41] I thank Mark Greenberg and A.J. Julius for pressing me to think about constitutivism carefully in terms of the chess analogy.

horizontally, that would not even count as a move in chess: to play chess, you would have to take that *attempt at a move* back, and make a legal move instead.[42]

In addition to constitutive rules, however, there are also further rules for *good* chess-playing, ones that do not seem constitutive of chess-playing as such. These are rules of *strategy*. For instance, it may be a good rule to castle in certain circumstances, or to avoid certain openings. And clearly I could play chess without following these rules, because I could play chess without having even *heard* of these rules, or otherwise learned them. Nonetheless, they are rules for good chess-playing. So there are non-constitutive rules for good chess-playing.

This is enough to motivate an accusation of equivocation. For it would be clearly false to say that I play chess well, *tout court*, whenever I manage to successfully make legal moves—whenever, that is, I manage to successfully follow the constitutive rules. What it is to play chess well in the sense of meeting all the norms that apply to chess-playing, and so in the sense of 'well' that makes the chess analogue of premise [2] true, is to successfully follow not only the constitutive rules, but also all the additional rules of good strategy. In contrast, if we try to substitute 'following only the constitutive rules' for 'playing chess' in the phrase 'playing chess well', then the sense of 'well' shifts. That new sense of 'well' makes reference, more narrowly, to following only the constitutive rules of chess successfully. And it is simply not true that to play chess well, *tout court*, it is enough to successfully follow the constitutive rules. Yet the substitution that forces the shift in the sense of 'well' would be required to make premise [2] interact with premise [1] in the way needed in a chess analogue of The Derivation: The Derivation trades on intersubstitutability. That is why, in terms of the chess analogy, the way that premise [2] interacts with [1] in The Derivation equivocates between two senses of 'well'.

The objection as applied to constitutivism about practical norms, of course, is that the accusation of equivocation also holds in the case of agency. Even if there are some constitutive rules, there may also be further, non-constitutive rules for good deliberation, just as there are non-constitutive rules for good

[42] Might there also be a constitutive "aim" of chess, namely, check-mating? I consider this complication below.

chess-playing. According to the objection, there are two different senses of 'well', a constitutive and a non-constitutive sense. One deliberates well in a constitutive sense whenever one manages to deliberate *successfully*—whenever one does fully deliberate. But surely, it seems, one could deliberate successfully in this sense while still deliberating *in a way one should not deliberate by some further standard*. In particular, one could successfully deliberate in accord with putative norms that one should not be deliberating in accord with in the first place. In the case of theoretical reasoning, it seems that one could reason successfully in accord with, say, counter-induction or affirming the consequent—rules that seem, intuitively, to be ones we should not follow. Likewise, the thought is, in the practical case you could deliberate successfully, without any performance errors, but in accord with bad rules of practical reasoning, whatever those might be, and whatever the further, non-constitutive standard for determining their badness. This would be a case of fully meeting the constitutive standards for deliberating, but failing to meet, or even to follow, further practical norms.

This objection admits that there may be some constitutive norms of agency, but objects that they need not be exhaustive of the practical norms there are, contrary to what The Derivation ambitiously claims. The accusation of equivocation therefore leads to the suggestion that there may be "hybrid" views of practical norms: views on which some practical norms are constitutive of agency, deriving their authority from the metaphysics of agency, while others may be non-constitutive norms, deriving their authority from some other possible source.

My response to this objection is two-fold. First, the chess analogy used to motivate it is faulty: in fact the norms of chess-playing are all constitutive, when properly understood. This leaves it open that the accusation of equivocation in the case of agency may stand on its own, if properly motivated. So second, I argue that, given the best reading of premise [1], there is simply no room, in the case of agency, for following putative non-constitutive rules of reasoning while also following the constitutive ones one must follow; and that hence there is no room for equivocation between two senses of 'well', either. The purported gap between senses of 'well' closes. This is enough to show that The Derivation as it stands is valid, given premise [1]. Objections to The Derivation cannot stand on their own, without seriously

engaging philosophy of action. Of course, no instance of The Derivation can stand on its own without seriously engaging philosophy of action, either. But constitutivists precisely do engage philosophy of action in thinking about the standards of practical reason. My point is that non-constitutivists need to do so too, to make room for their views.

This does not yet rule out the thought that such an engagement with the philosophy of action could yield a weaker premise than [1], and a correspondingly weaker argumentative schema that vindicates *some* constitutive norms while also leaving room for non-constitutive norms. But I outline some difficulties for making sense of such "hybrid" views.

1.3.2.1 First response: against the chess analogy

The first response to the accusation of equivocation is that all norms of chess are, in a sense, constitutive norms of chess; and that therefore the chess analogy cannot motivate the accusation. However, my first observation looks to tend in the opposite direction. The observation is that the putative constitutive norms of chess, the rules defining legal moves, are inviolable, and so cannot be normative at all. There is simply no room for cases of moving the bishop as a move in chess while, say, moving the bishop three squares diagonally and then one square horizontally. You either follow the rules defining legal moves perfectly successfully, or you do not follow them at all. But again, where there is no room for following a rule badly, there is really no sense to be given to following it "well," either. Violability is a condition of normativity.[43]

Of course, all that this means is that there is no such thing as following the rules definitive of legal moves badly *by the lights of those very rules.* There is still room for following them badly by the lights of further standards such as the standards of strategy. (In following standards of strategy, one must

[43] Notice that it is not enough for violability in the relevant sense that the rule itself *rules certain things out*. The rule for moving the bishop does rule certain things out: for example, moving the bishop horizontally. But this bare idea of "ruling out" is distinct from, because weaker than, the idea of its being possible to follow a rule well or badly while still being engaged in the activity of following it at all. It is the latter, stronger criterion of violability that the rules definitive of legal rules in chess fail to meet. And that is what makes those rules incapable of being normative, as opposed to merely descriptive of what it is to play chess at all.

also be making legal moves.) Are these further standards non-constitutive or constitutive? We previously classified them as non-constitutive. But in fact this seems wrong. For like most games, chess also has a constitutive *aim*, by achieving which one wins: check-mating. One would not be playing chess unless one made legal moves with this aim.[44] Since this aim is very well-defined, it makes good sense to think that the rules of strategy simply derive from this aim, when taken together with the rules definitive of legal moves. Some ingenuity must go into coming up with the rules of strategy, to be sure: the "derivation" is not mechanical. But nonetheless, there is a sense in which the rules of strategy simply flow from what it is to effectively pursue the constitutive aim within the constraints of the legal moves available.

Does this mean that, since only the rules of good strategy are such that, in following them, one plays chess "well," there are only constitutive norms for chess? There is a complication: as we previously noted, it seems that one *can* play chess without having learned, and so without following, any of the good rules of strategy. One might play with the aim of check-mating while making nothing but bad moves. So it is *not* constitutive of playing chess to follow the rules definitive of good strategy. Are there, therefore, no constitutive norms for chess-playing, despite its constitutive aim?

Well, there is at least *one* constitutive norm, which any chess-player must follow, and can follow well or badly: "Check-mate your opponent!" Together with the rules defining legal moves, this gives rise to further norms, those of good strategy. These norms are not constitutive in the *strict* sense in which one must follow them to be playing chess at all. But they are nonetheless derivative from the constitutive facts about chess. And this makes them "constitutive" in an *extended* sense that is certainly of interest to the project of grounding norms: the rules of good strategy derive from the constitutive facts about chess, and relatedly, we can see them to be good rather than bad strategic rules by reference to those same constitutive facts about chess. Correct views about which rules of strategy are good ones can therefore be shown to be well-grounded simply by reference to what chess is.

---

[44] Or, as Larry Crocker points out to me, the aim of achieving a stalemate. Apparently, many Grandmaster games are such that the person who ends up playing black can realistically only aim at a stalemate.

If something similar could be shown to hold in the case of agency and practical norms, this would amount to the sort of normatively non-question-begging grounding for practical norms that constitutivism seeks to give. The present point, however, is just that these reflections on chess undermine the case for the alleged equivocation between two senses of 'well'. The only norms of chess there are are the norms for good strategy, and they are constitutive norms. (One is a constitutive norm in the strict sense and several are constitutive norms in the extended sense.) If agency has a constitutive aim, similar points might apply. We need some further motivation for the accusation of equivocation, if it is to stick.

An idea that is present in these remarks is that the notion of a *telos* could help after all, in helping us frame the one constitutive norm that figures in The Derivation. I return to this in §1.3.2.2. (Of course, this would not undermine the points in §1.2.2, namely, that it matters what the *telos* is and how one may "aim" at it.)

1.3.2.2 Second response: no room for non-constitutive rules of reasoning given constitutive rules

I outlined the second response above by saying that, given the best reading of premise [1], there is simply no room, in the case of agency, for following putative non-constitutive rules of reasoning while also following the constitutive ones one must follow. This, I said, would show that there is no room for equivocating on 'well'; and further, that unless all instances of [1] are false, constitutivism has a monopoly on practical norms.

To see why these claims hold, start by considering what [1] says, and what practical reasoning, or deliberation, is. [1] says that to deliberate *just is* to reason in accord with R, and in this sense to follow R. If to deliberate just is to follow R, then one could not deliberate without reasoning in accord with R. But now, this does not appear to leave room for reasoning in accord with other, non-constitutive rules. For (perhaps unlike rules of chess), full-fledged rules of reasoning seem *exclusive* of each other. What is a full-fledged rule, and what do I mean by exclusivity?

To be a full-fledged rule of reasoning, a rule R must be such that, in following it, one is, fully, reasoning: one cannot fall short of reasoning in following R, if R is a full-fledged rule of reasoning. R in

premise [1] is a full-fledged rule of reasoning, since to follow it just is to deliberate. (Another way of being a full-fledged rule of reasoning would be being just one among many possible ways to deliberate.) I will say more about being "full-fledged" below.

Cases of theoretical reasoning help to illustrate what I mean by exclusivity. When you reason in accord with modus ponens, it seems that you cannot, in that very same reasoning, also be reasoning inductively or abductively, nor vice versa.[45] Likewise, if you reason in accord with modus ponens, then you cannot, in that very same reasoning, be reasoning in accord with, say, affirming the consequent. There may of course be arguments or strings of reasoning that incorporate, as stages, reasoning in accord with one full-fledged rule as well as reasoning in accord with another. (Such stages may even be executed simultaneously in a reasoning system: timing is not the point.) Nonetheless, you cannot literally reason in accord with two different full-fledged rules in the same reasoning. Why? For one thing, different rules of reasoning (often) take different things as premises. For another, rules of reasoning describe transitions between premises and conclusions. If there is a different rule of reasoning, then there is a different transition—and correspondingly, either different premises, a different conclusion, or both. And you cannot be making two different transitions, take two different sets of premises as your premises, or draw two different conclusions, in the very same piece of reasoning.

If rules such as modus ponens and induction are full-fledged rules of reasoning, then, they seem to exclude each other in the sense that one cannot, in the very same reasoning, be reasoning in accord with both. And modus ponens and induction do seem to be full-fledged rules of reasoning. In following modus ponens, one reasons. In reasoning inductively, one reasons. But now suppose a constitutivist about modus ponens claims that it is a metaphysical fact about reasoning that one must reason in accord with modus ponens to reason at all. Then one cannot reason inductively at all. Indeed, one cannot reason in accord with *any* other rule besides modus ponens. (Proof: Suppose, for reductio, that one reasons (say) inductively; then one reasons; which implies (according to the constitutivist about modus ponens) that

[45] While logical principles such as modus ponens only state facts about implication, I will assume that we can easily formulate putative principles of reasoning roughly corresponding to those facts.

one reasons in accord with modus ponens; which, given exclusivity, in turn contradicts the initial assumption that one reasons inductively.)

Suppose, then, that some practical rule R is a full-fledged rule of reasoning, in the way in which modus ponens and induction seem to be full-fledged. And suppose, as per premise [1], that to deliberate at all just is to reason in accord with R. Then it looks like there simply cannot be two different rules of reasoning that are genuinely authoritative, but one of which is non-constitutive: the constitutive rule R is the only authoritative rule governing how one should deliberate. This follows by a very plausible, because a very weak, "ought" implies "can" principle:

> **"Ought" implies "can" principle for rules of reasoning:**
> If you ought to reason as a rule of reasoning P prescribes, then it must be the case that it is possible for some rational agent or other to reason in accord with P.[46]

This principle looks true. What could possibly be the point of saying that we ought to follow some non-constitutive rule, if the very nature of rational agency makes following it flat-out impossible?

I have been presenting a quick argument to the effect that, if some version of [1] is true, then there is no room for non-constitutive standards of practical reason. In terms of the charge of equivocation, we can put this point by saying that there is no room for the charge of equivocation to get a grip, since there could not be more to good deliberation and acting for reasons than following the constitutive rule R well. In sum, the argument for this conclusion is this: For [1] to be true, R must be a "full-fledged" rule of reasoning, in the sense that one cannot fall short of deliberating in reasoning in accord with R; but if R is both full-fledged in this sense and constitutive of reasoning as such, then there is no room for also reasoning in accord with non-constitutive rules. Given the weak "ought" implies "can" principle, no non-constitutive rules could be authoritative. Unless we can challenge all versions of [1], constitutive norms

[46] Notice that the principle is much weaker than "internalism" about normative reasons for action. According to internalism, *p* cannot be a reason for some individual agent A to φ unless *A herself* is capable of being moved to φ by the belief that *p*.

are all the practical norms there could be. Call this the Exclusivity Argument. Let me consider objections to it, including possible weakenings of [1].

Objection A stems from a reminder about where we started at the outset to §3. Korsgaard's constitutivist theory seemed to encompass *two* different constitutive "rules" of reasoning, a categorical imperative CI and a hypothetical imperative HI. In Korsgaard's theory, recall, one follows HI to formulate a maxim, and following CI gets one to actually act on this maxim (see pp. 28-29 above). If CI and HI are full-fledged constitutive rules of reasoning, then mustn't there be something wrong with the Exclusivity Argument? For if the Exclusivity Argument implies that there could be no non-constitutive norms of reasoning, then it also looks to imply that there could be only *one* full-fledged constitutive rule. And quite generally, abstracting from the details of Korsgaard's view, we might be suspicious of the conclusion that constitutivists must hold that there is only one constitutive norm of deliberation. Yet that is what the Exclusivity Argument looks to entail.

In response, note that CI and HI are not in fact full-fledged rules of reasoning in Korsgaard's theory. For recall that following just *one* of CI or HI (or neither) is not enough for deliberation, on that theory: one must follow both to deliberate at all. They are individually necessary, individually insufficient but jointly sufficient conditions of deliberation. But if "following" just one of these rules on its own is not enough for deliberation, then whatever "following," say, HI on its own might mean, it cannot mean "reasoning in accord with HI." Correspondingly, then, neither can HI be a norm telling us how we *should* deliberate. It cannot be an authoritative rule of reasoning on its own simply because it cannot be a rule of reasoning on its own. Likewise for CI. For Korsgaard, there are not two constitutive norms of deliberation, but one: R = (CI+HI).

We might of course say that CI and HI are *aspects* or *parts* of deliberation as such, and so aspects or parts of good deliberation as such.[47] But to be an aspect or part of deliberation in this sense is not the same as being a stage in an argument, as a piece of reasoning by induction might be a stage in an

[47] This is in fact what Korsgaard says, although she prefers to put things by saying that HI is an aspect of CI, not that they are each aspects of deliberation just as such (2009: 70).

argument that also incorporates different pieces of reasoning by, say, modus ponens. So nothing in the "aspect" thought speaks against the Exclusivity Argument.

The same points hold more generally, not just for Korsgaard's view. If a rule is not full-fledged, then it is not a rule of reasoning. It follows that, if the putative constitutive norm R in premise [1] is a full-fledged rule of reasoning akin to rules like modus ponens, as we have been supposing that it is, then the Exclusivity Argument goes through (barring further objections). One implication of this is, indeed, that if practical norms must be full-fledged rules of reasoning, then just as there is no room for non-constitutive practical norms, likewise there is only room for one constitutive practical norm. But in the end, this should not be too surprising. For suppose, for reductio, that there are two full-fledged rules of reasoning that are each constitutive of deliberation just as such. Since each is full-fledged, one can deliberate fully by following either. But then one can deliberate fully without following the one, and one can deliberate fully without following the other. Hence neither is such that following it is metaphysically necessary for deliberation just as such. So neither is a constitutive norm, contradicting our supposition that there are two.

This line of response to Objection A raises further objections. One type of objection is to the assumption that practical norms must be "full-fledged" rules of reasoning. This amounts to a weakening of [1]. Another type of objection is to the Exclusivity Argument on different grounds: it charges that even if constitutive norms are full-fledged rules of reasoning, in the sense that in following them one fully reasons, these rules might nonetheless be too indeterminate to rule out one's following other, perhaps even non-constitutive, norms. I take the latter type of objection first.

Objection B takes a cue from our discussion of chess in §1.3.2.1.[48] Recall that there is only one norm of chess that is constitutive in the strict sense in which one does not count as playing chess at all unless one follows that norm. That norm was "Check-mate your opponent!" But there are many different ways one might try to check-mate one's opponent within the confines of the legal moves, depending on

[48] Kieran Setiya put something like this objection to me. I hope that, in developing it, I have not distorted it too badly.

one's situation and on the opponent's moves. These different ways may be thought of as different rules of strategy: for each move, we may try to determine whether it was a good move or a bad move in relation to the aim of check-mating, and so whether it was strategically a good move or not. And now, recall that one can follow many different rules of strategy, not all of which, perhaps even none of which, need be good. The good rules of strategy are constitutive in the extended sense that they are ways of following the central norm of check-mating the opponent, and that their goodness is explained by this relation to that central norm. But *bad* rules of strategy do *not* stem from the central norm of check-mating the opponent—at least insofar as they are bad. Yet one can follow them, if one is a bad chess-player. So one can follow non-constitutive rules even while playing chess, and so, even while following the strict constitutive norm "Check-mate your opponent!" By analogy, the objection suggests, we might be able to follow very open-ended constitutive rules such as "Maximize knowledge!" while following other rules, including, possibly, non-constitutive ones.

In response, we might worry about how we could even think of "Maximize knowledge!" as a rule of reasoning. What would it be to reason in accord with it? But for present purposes, it is enough to consider it a codification of the singular tendency that, according to the relevant version of premise [1], any sub-type of practical reasoning must have in order to be a type of practical reasoning. That is, to deliberate at all one must reason in some way that, in one's circumstances, would maximize knowledge if one reasoned in that way fully successfully. However, if something like this version of [1] is true, then it is hard to see how it *is* possible to follow rules of reasoning that are not constitutive, at least in the extended sense introduced in §1.3.2.1. For all the sub-rules we can follow that are rules of reasoning according to the relevant version of [1] are ways of maximizing knowledge in the situation one is in. They owe their status as sub-rules of reasoning to the alleged strictly constitutive fact that all reasoning just as such seeks to maximize knowledge. And if they are good sub-rules to follow, they can be seen to be so by reference to this strict constitutive fact. Thus even though we may think of some constitutive norms in the strict sense as being very open-ended, nonetheless under the supposition that the corresponding version of premise [1] is true, there is no room for rules of reasoning that are not constitutive at least in the extended

sense. (Recall that we are still operating under the assumption that rules of reasoning are full-fledged.) Hence, given the "ought" implies "can" principle for rules of reasoning, neither could any putative non-constitutive rule of reasoning be authoritative.

In any case, even in the case of chess, it is hard to see how any putatively non-constitutive rule could be a *good* rule for playing chess, given that good rules of strategy derive from the strict constitutive norm "Check-mate your opponent!" What good non-constitutive norms of chess might there be left? One might propose that e.g. added time-constraints, such as those in speed-chess, are not constitutive norms of chess-playing, but neither do they contradict any constitutive norms of chess; still, they are certainly norms that one should observe in playing speed-chess.[49] However, such added norms are precisely *norms of speed-chess*—and constitutive norms of speed-chess, at that—not norms of chess, as such.[50]

This line of response to Objection B concedes that there is something wrong with the Exclusivity Argument, as it was quickly formulated at the outset to §1.3.2.2. What is wrong with that quick formulation is the supposition that there couldn't be many different sub-rules that are each rules for full-fledged reasoning, but that one nonetheless need not follow in any exercise of agency just as such. However, a more sophisticated version of the Exclusivity Argument still goes through. If some version of premise [1] of The Derivation is true, as we have been supposing, then we must interpret any sub-type of reasoning as a way of following, in one's situation, the central rule R in terms of which [1] is stated. And as we saw, this still leaves no room for reasoning in accord with putative non-constitutive rules, and so, no room for such putative non-constitutive rules to be authoritative. Thus, even if [1] is formulated in terms of open-ended rules, constitutive norms are all the practical norms there could be.

As far as I can see, there are no other relevant objections to the Exclusivity Argument, or indeed to The Derivation itself, that are not also objections to [1]. The Exclusivity Argument rebuts the accusation of equivocation by showing that there could be no room for such equivocation, given [1] itself:

---

[49] Thanks to Kieran Setiya for suggesting this.

[50] Of course, it is optional, within a certain range, exactly what time constraints one sets. Nonetheless, all time constraints observed are ways of observing the central constitutive norm of speed-chess we might roughly formulate as "Check-mate your opponent, but move fast!"

if there are constitutive rules of reasoning of the sort that figure in [1]—rules such that to deliberate *just is* to follow them (in some way or other)—then there is simply no room for extra, non-constitutive norms that could fuel the charge of equivocation. Thus if some version of [1] is true, constitutivism does indeed have a monopoly on practical norms. Our only options are to either embrace constitutivism in some version, or else show that all versions of [1] are false.

One could concede all of this, but try to object to [1] itself while nonetheless holding out hopes for some constitutive standards of practical reason or other. In particular, perhaps practical norms need not be "full-fledged" rules of reasoning of the sort that figure in [1], but can instead govern mere "aspects" of reasoning. On this suggestion, the first, metaphysical premise of a derivation of such "aspectual" constitutive norms R should read something like:

1*. To deliberate, it is necessary that one follows R, in some sense of 'follows'.

We might be able somehow to derive the authority of R from this, making it a constitutive norm; and yet on this permissive proposal, there might also be room for non-constitutive norms. Call this *Objection C.*

In response, however, it is hard to see how the derivation might go, or what the proposed premise [1*] even means. We already noted that "following" R cannot be a matter of reasoning in accord with R, if R is supposed to govern only an aspect of reasoning. But nor could "following" R be a matter of mere external conformity to R's dictates: for we constantly externally conform to an infinity of possible rules we could formulate, and it would be absurd to say that we thereby "follow" them—certainly in any sense of "following" that could further imply those rules' authority on us. One might suggest that the sense in which one "follows" an aspectual norm R whenever one deliberates is that R picks out a disposition that necessarily manifests itself to at least some extent whenever one deliberates. But again, if this disposition is not a disposition of reasoning, then why should we think that it corresponds to a standard of practical

reason, any more than (say) the disposition to take some time while deliberating corresponds to a standard of practical reason?[51]

There are also independent problems with how to mount a derivation of the authority of the purported aspectual norm, even supposing we can make adequate sense of "following" it. We cannot simply plug in [1*] to The Derivation: in this case, the accusation of equivocation really *would* be fatal, and no helpful analogue of [4] would hold. More simply, if the standards of practical reason correspond to deliberating well, *tout court*, then why hold that a mere aspect of deliberation corresponds to any standard of practical reason at all?

Finally, consider the consequences if there are aspectual constitutive norms that are, in and of themselves, genuine standards of practical reason. Suppose that R is such an aspectual constitutive norm. Then either there are further constitutive norms that correspond to all the other aspects of deliberation, or there are not. If there are, then R and the other purported aspectual constitutive norms can simply be plugged into [1] in The Derivation, like Korsgaard's (CI+HI). There is no genuine alternative here to [1]. If there are no further aspectual constitutive norms, however, then either there are no further genuinely authoritative practical norms at all, or there are non-constitutive practical norms. If there are no further genuinely authoritative practical norms, then aspectual constitutive norms are all the practical norms there are. This amounts to a skepticism that there is such a thing as deliberating well *tout court* at all. While such skepticism might be true for all I have said here, it certainly seems a drastic move to derive such skepticism from the tenuous idea of aspectual constitutive norms. A more promising proposal might be to go in for a "hybrid" view, on which there are some aspectual constitutive norms as well as some aspectual—or perhaps even full-fledged—non-constitutive norms. However, there are general problems affecting the very idea of "hybrid" views about the standards of practical reason. These problems arise whether the non-constitutive and constitutive norms we try to combine are aspectual or full-fledged.

---

[51] Possible answers: Perhaps the disposition to take some time while deliberating is not a necessary aspect of deliberation. Or perhaps it is, but it is impossible to fail to take some time.

What problems? A hybrid view is committed to holding that there are two quite different sources of authority for practical norms, depending on the norm in question. One type of norm derives from the metaphysics of agency, the other from something else. Yet each of the two sorts of practical norm is supposed to govern the same activity, or aspects of the same activity—that of deliberating and acting for reasons. This sort of view is problematic. The problem is not just that positing such a deep bifurcation among sources of practical norms offends a certain philosophical sensibility and a search for unity. There is something to be said for unity, but I doubt it can be a decisive consideration on its own. The trouble is, rather, that were we to posit such a bifurcation among the sources of practical norms, we would also need an *explanation* of how it is that two sets of practical norms that have entirely distinct sources nonetheless manage not to conflict with each others' dictates. Or: could the realm of practical norms, taken as a whole, be self-contradictory? That would be more surprising still.

One might think that this problem does not arise in the case where there are only aspectual constitutive norms and aspectual non-constitutive norms: we could try to mount an "Aspectual Exclusivity Argument" to show why not, to the effect that given that an aspect of deliberation is governed by a constitutive norm R, that same aspect cannot also be governed by some different, non-constitutive norm. However, it is hard to see how to fund the exclusivity claim needed for such an argument. Full-fledged rules of reasoning excluded each other in the sense that one could not reason in accord with one full-fledged rule in the very same reasoning while reasoning in accord with another. This ruled out hybrid views of full-fledged constitutive norms and full-fledged non-constitutive norms (quite in advance of the "two sources" worry), because the idea that one might follow a full-fledged non-constitutive rule of reasoning simply negates the idea that, to reason at all, one must follow a different, constitutive rule, however imperfectly. In contrast, suppose that, as on the interpretation that looks most promising, "following" an aspectual constitutive norm is a matter of manifesting a disposition of some sort—but not, of course, a disposition of full-fledged reasoning. It is hard to see how anything in the idea of a disposition, even a metaphysically necessary one, yet conflicts with the thought that there might be non-constitutive norms requiring you to repress it or resist its manifestation, as much as possible. To make the

possibility palpable, just think of puritanical commands to repress dispositions necessary for *human* nature, as such, or for *this* or *that* human, as such. (At least in the case of dispositions necessary for individual humans, the call for repression need not be particularly puritanical to be imaginable.) Such calls for repression might be mistaken, but it is hard to see how we could derive this conclusion from the mere idea of a disposition necessary for the type of being of whom the repression is demanded. If the claim of exclusivity cannot be funded, however, then the "two sources" problem remains. If the two sources of practical norms are completely independent of each other, what guarantees that different putatively authoritative practical norms will not conflict with each others' dictates?

I have been responding to Objection C by presenting problems with the idea of "aspectual" standards of practical reason, whether in a "hybrid" or "non-hybrid" guise. If these problems are soluble, then the aspectual idea may provide one viable way of resisting premise [1] of The Derivation. Regardless, however, we are faced with the stark options I presented above: either embrace some version of full-fledged constitutivism, or else show that all versions of [1] are false.

1.4 CONCLUSION

In this chapter, I argued, first, that "function" arguments do not help explain the constitutivist derivation of "ought" from "is"; and second, that there is a quite powerful argument schema that does explain the derivation. According to this argument, if there are any (full-fledged) constitutive norms at all, then they exhaust the realm of practical norms: there is simply no room for non-constitutive norms of deliberation and acting for reasons. This yields the result advertised at the outset. Our only options are either to embrace constitutivism, or to give an account of agency on which there is no "R" that makes premise [1] of The Derivation true. In either case, anyone who cares about what the practical norms governing agency are cannot afford to ignore some rather in-depth questions in the philosophy of action. Philosophy of action is inescapable for metaethics.

As I mentioned at the outset, my own view is ultimately anti-constitutivist. Although we can explain how the derivation of "ought" from "is" works, given a certain sort of "is"-claim, I doubt that any "is"-claim of the relevant sort is true. I argue for this claim in chapters 2-3. But we can briefly gesture at the relevant anti-constitutivist thought in intuitive terms. I said earlier that, whenever you "follow" a full-fledged rule of reasoning, you deliberate in accord with it. And when you deliberate in accord with some R, you deliberate. I also raised the examples of induction, modus ponens, abduction, and so on. These all seem to be full-fledged rules of reasoning. So if you reason inductively, you reason. And if you reason in accordance with modus ponens, you reason. These disparate possibilities for full-fledged theoretical reasoning make it seem like there might be no constitutive rules of theoretical reasoning. Although it is true that, whenever one reasons, one must always reason in some way or other, in accord with some rule or other, it is not the case that there is a rule such that one must always reason in accord with *it* in order to reason at all. What my dispositions of reasoning are may be a contingent fact about me, not a necessary fact about theoretical reasoners as such. Still, as long as those dispositions are intelligible as dispositions of reasoning, I am intelligible as a reasoner. The anti-constitutivist thought I defend in chapters 2-3, with regard to practical reasoning and acting for reasons, is just this same thought, applied to dispositions of practical reasoning.

If that anti-constitutivist thought is right, then the nature of agency cannot afford a normatively non-question-begging defense of any practical norms. This leaves us with the problem described at the very beginning of §1.1: unless we can give some other normatively non-question-begging defense of our specific normative commitments, we must face the question whether and why our specific normative commitments are anything more than arbitrary declarations of allegiance to a way of life.

2.0 INSTRUMENTALISM AND PRACTICAL RULE-FOLLOWING

2.1 THE QUESTION: IS INSTRUMENTALISM IMPLICIT IN THE CONDITIONS OF AGENCY?

According to instrumentalism, all normative reasons for action take an instrumental form. Roughly, one has normative reason to φ if and only if, and because, φ-ing is a means (on some conceptions, the most efficient or a necessary means) to some end ψ one antecedently desires or intends. Correspondingly, good practical reasoning is good instrumental reasoning. It is, in a sense to be explored below, reasoning in accord with an "instrumental principle" that tells one to pursue believed means to one's ends. Provided one's beliefs about means are correct, one who reasons instrumentally well also acts as she should, and for good reasons.

Though sometimes suspected of incoherence, and often reviled for its counterintuitive consequences—for example, that one has reason to refrain from harming others only if one happens to want something that refraining helps one get—instrumentalism continues to attract.[52] Even anti-instrumentalists usually accept that there are *some* instrumental standards of practical reason, seeking only to supplement these with non-instrumental ones.[53] Some recent authors have diagnosed instrumentalism's enduring attraction as due to an inchoate appreciation of a deep truth about the nature of agency: that

[52] For the charge of incoherence, see Korsgaard 1997; Quinn 1993 articulates similar concerns from a different perspective.
[53] This tendency is documented in e.g. Vogler 2002, Introduction and ch.1. A notable exception to the tendency is Raz 2005a, b.

unless one heeds instrumental standards to at least some extent, one cannot act for reasons at all.[54] While the diagnosis may be correct, I think the attraction therein diagnosed is mistaken. The nature of agency does not support instrumentalism, nor does it support the weaker idea that instrumental standards have an especially secure status among putative standards of practical reason. My aim in this chapter is to show why, and thereby to dispel the attraction.

The attempt to support instrumentalism by appeal to the nature of agency is an instance of the general strategy of *constitutivism*. Constitutivists argue that (a) exercising agency just is a matter of following some specific rule(s) or norm(s) R, and further, (b) all and only the R that are in this way constitutive of agency authoritatively specify what good practical reasoning and reasons amount to. Indeed, R's normative authority is explained by R's being constitutive of agency: since agency as such just is a matter of following R, following R *perfectly* must be what it is to be good *qua* agent, and following R imperfectly or badly must amount to being a bad agent. A good agent is just someone who does well what agents as such do—namely, the activities of deliberation and acting for reasons. Practical norms are explained in terms of proper psychological functioning.[55]

If R is both constitutive of agency and authoritative because constitutive, let us say that it is a "constitutive norm" of agency. Depending on the content of R, constitutivism can be used to support a variety of views about what the standards of practical reason are. Both instrumentalists and anti-instrumentalists have recently mounted constitutivist defenses of their views: for example, James Dreier (1997, 2001) and Candace Vogler (2002) of instrumentalists,[56] and Christine Korsgaard (2009), David Velleman (2009) and, more tentatively, Michael Smith (2009, 2010), of anti-instrumentalists. In arguing against constitutivist instrumentalism, my target is its premise, (a). I argue that there can be non-

---

[54] E.g. Dreier 1997, 2001; Vogler 2002, esp. Introduction, chs.1, 6 & 7; Smith 2009, 2010.

[55] Korsgaard 2009: 27-47 and Smith 2010: 125 both put their constitutivisms in terms of the vocabulary of "proper functioning." Smith 2010 also emphasizes the metaphysical modesty of constitutivist accounts of normativity as a point in their favor.

[56] Vogler 2002 calls her view "calculative", to distance herself from the moral-psychological commitments of many contemporary instrumentalists. Nonetheless, she understands her view as capturing the hidden core truth in contemporary instrumentalism (Introduction and ch.2, 26-29).

instrumental agents, agents who heed no distinctively instrumental standards of practical reason at all. However we might support the constitutivist's inference from a metaphysical "is"-claim, about the nature of agency, to an "ought"-claim, about good practical reasoning and reasons, constitutivist instrumentalism cannot get off the ground because its premise is false. (My argument will thus also undermine, in part, those constitutivist anti-instrumentalists, such as Korsgaard and Smith, who allow that an instrumental norm is *a* constitutive norm of agency, even if only along with other, non-instrumental norms.)

While my target is constitutivist instrumentalism in all its forms, I take as my initial focus James Dreier's argument, in his (1997) "Humean Doubts about the Practical Justification of Morality." Though Dreier is a self-described "Humean," his argument proceeds on the basis of regress considerations, reminiscent of Lewis Carroll's in "What the Tortoise Said to Achilles," that at least initially look to have perfectly general appeal. I think reflection on this regress can yield genuine insight into the nature of agency—though not, as I argue, in the way Dreier proposes. Dreier argues that the regress considerations reveal reasoning in accordance with an instrumental rule to be a condition of agency. In contrast, I argue that reflection on Dreier's regress helps to show the opposite: there can be non-instrumental agents, agents who reason only in accordance with other, non-instrumental rules, and correspondingly, act only for non-instrumental reasons.

In fact, my argument will tentatively suggest an even stronger conclusion, of interest to all constitutivists: that agency is a general capacity for practical rule-following, correctly attributable to anyone who is intelligible as following some rule among many possible candidates; but that there is no unique rule that agents as such must follow to be agents. If this is right, then there are no "constitutive norms" of agency. The conditions of agency are compatible with following any among many non-overlapping subsets of putative practical rules.[57] The result is a "content skepticism," not about practical reason,[58] but about constitutivism: about the extent to which insights into the content and structure of the

[57] As will become clear, talk of "rules" is to be taken loosely.
[58] Korsgaard 1986b introduced the term "content skepticism" about practical reason, contrasting it to "motivational skepticism" about practical reason and arguing that the latter depends on the former.

domain of standards of practical reason can come in the form of claims about the nature of agency. Though this general suggestion remains tentative in this chapter, it is worth highlighting. For if it is right, it threatens all constitutivisms in practical reason. Not only can instrumentalism not be based on the conditions of agency, but neither can any legitimate anti-instrumentalism.

§2.2 outlines Dreier's argument and raises initial difficulties for it. §2.3 considers how the argument might be strengthened; in seeing how it fails, we find support for the possibility of non-instrumental agency, as well as for content skepticism about constitutivism (§§2.3-2.4).

Note that since my target is constitutivist instrumentalism, I will not here engage non-constitutivist arguments for instrumentalism, such as those based on non-constitutivist analyses of the concept of a normative reason, or those reliant on first-order normative intuitions.[59] Even if my argument against constitutivist instrumentalism stands, instrumentalism could be true on other grounds. That said, constitutivist instrumentalism is worth engaging on its own. This is in part because constitutivism in general is so ambitious. The constitutivist project promises to provide almost everything one could want from a complete practical philosophy: an account of agency, an identification of authoritative practical norms, and an explanation of the authority of those norms—and all this without relying on first-order normative intuitions that a skeptic could deny.[60] But furthermore, out of all putative constitutive norms of agency, an instrumental norm of reasoning is widely taken to be the *least* controversial. It will be worth seeing, then, what aspects of agency may have misled us into believing that an instrumental norm is implicit in agency. As a result, we also achieve an independent payoff: an attractive new conception of

---

[59] Schroeder 2007a,b argues for "Hypotheticalism," a view akin to instrumentalism, on the basis of a non-constitutivist analysis of the notion of a normative reason together with normative intuitions and broad methodological considerations. The classic argument for instrumentalism is based on "internalism about reasons" (the view that agents must be capable of acting on the reasons that are authoritative for them), together with the "Humean" theory of motivation. This argument is discussed in e.g. Williams 1979, Korsgaard 1986b, Hooker 1987, Wedgwood 2002. Constitutivist arguments are stronger than each of these types of argument: constitutivism entails internalism without presupposing it, and does not rely on normative intuitions. Nonetheless, since I will be, in part, arguing against the Humean theory of motivation, my argument will also tell against internalist arguments for instrumentalism.

[60] Dreier 1997, 2001 and Korsgaard 1996, 2009 both emphasize the way in which constitutivism promises to ground practical norms in normatively non-question-begging terms.

agency as a general capacity for practical rule-following, a capacity in which no commitment to a specifically instrumental norm is implicit.

2.2 DREIER'S REGRESS: THE M/E SKEPTIC

James Dreier (1997) mounts a regress argument purporting to show that the following instrumental rule M/E states a constitutive norm of agency:

> **M/E** [For all φ and all ψ] If you desire to ψ and believe that by φ-ing you will ψ, then you have a reason to φ.[61]

The regress arises as follows. Let a sincere M/E-skeptic be a putative agent who does not "accept" M/E in the sense that she is not disposed to actually follow it: she is not disposed to act on the instrumental reasons that M/E picks out as applying to her (89).[62] Now suppose such a skeptic desires an end ψ and believes that by φ-ing she will ψ, but she is not disposed thereby to φ. And suppose we tell her, citing M/E, that given her beliefs and desires, she has reason to φ; so barring countervailing reasons, she should φ. But she fails to find the proffered instrumental reason gripping precisely because she rejects M/E. So she asks for a practical justification of M/E itself: "Why exactly should I φ given my beliefs and desires? I already believe that M/E recommends that I do so, but why should I follow M/E? Give me a reason to follow it. If I am convinced, I will do so." What practical justification of M/E could we offer to the skeptic, so that she could, by appreciating the force of our purported justification, come to "accept" M/E?

---

[61] This is Dreier's formulation in his 2001 revised version of 1997, except that I add the universal quantifiers, which I think are intended in Dreier's discussion. (All page references are to the 1997 version unless otherwise specified.) I assume instrumental reasons can be defeated by other instrumental reasons.

[62] We might wonder why we should construe a skeptic about a rule as lacking a motivational disposition. It seems we could doubt a rule's authority even as we continue to be disposed to follow the rule (cf. Williams 1985: 28). But for Dreier, the regress is supposed to be a device for investigating the conditions of agency: this is why the motivational disposition is central.

Noting that it is a "kind of methodological axiom that whatever is missing when someone has a belief and lacks a certain motivation is a desire of some sort or other" (93), Dreier considers the suggestion that we could get the M/E-skeptic to "accept" M/E by getting her to desire to follow M/E. However, Dreier argues, this could not possibly work. For it would be to add a desire *for a further end*; and the skeptic is precisely someone who is not moved to pursue acknowledged means to her desired ends (93). Getting an agent who believes that by φ-ing she will ψ to desire to ψ is "a good way to motivate normal, rational agents" to φ, but in the M/E-skeptic's case it is "futile" (93). Nor will it help to retreat to second-order desires to have one's desire to follow M/E be effective: such second-order desires would not themselves be effective in one who fails to be appropriately moved by her desires for ends. A regress threatens. Insofar as successful practical justifications of M/E must get the skeptic to follow M/E, they must motivate her to do so, and this seems to imply that we must get the skeptic to desire to follow M/E; but it appears that no such desire *can* ever actually get the skeptic to follow M/E, even if we could persuade her to have the desire.[63] Thus Dreier concludes that the skeptic's requests for reasons to follow M/E are ultimately "empty" (96), or do not "make sense" (98). The M/E-skeptic is stuck in a regress of putative practical justifications of M/E, none of which she can heed precisely because she rejects M/E itself.

Now, Dreier conceives of this regress as analogous to Lewis Carroll's famous regress of justifications for *modus ponens* (MP). Just as Achilles cannot give the Tortoise a justification of MP that the Tortoise can heed unless the Tortoise already implicitly accepts MP—"accepts" in the sense of reasoning in accord with it, not merely in the sense of believing successive statements of MP as a premise—likewise Dreier thinks we cannot give the M/E-skeptic a justification of M/E that she can heed unless she already implicitly "accepts" M/E (Dreier 1997: 94-5; Carroll 1995). And this might make us

[63] Note that the argument does not thus far assume the "Humean" thesis that the only practical justifications that could rationally persuade one to adopt the desire to follow M/E must make at least implicit appeal to some further desire. Dreier's argument proceeds downstream from the adoption of desires: the idea is that desires, however adopted, can only move one to act given a disposition of reasoning, in particular, acceptance of M/E. (Cf. Nagel 1970: VI 1-2, Vogler 2002: 18-21.) §2.3 below considers how appeal to the Humean theory of motivation might strengthen the argument.

suspect that the regress will likewise arise for other rules R. For what appears to generate the regress is the claim that the M/E-skeptic lacks acceptance of exactly the rule of reasoning with which, Dreier claims, any putative justification of M/E would have to engage to motivate her. But likewise, it seems, neither could a skeptic about some other rule R be made to reason in accord with R merely by offering her justifications of R that depend for their efficacy on one's already accepting R.

Dreier argues, however, that no similar regress threatens with respect to other rules R, precisely because we can offer an R-skeptic justifications to follow R that do *not* depend for their efficacy on one's already accepting R. One who doubts some R distinct from M/E can in principle be got to follow R, *so long as she accepts M/E* (94). If we can get an R-skeptic who accepts M/E to desire to follow R, this will be enough to get her to follow R, given the general "methodological axiom," and provided she also knows the means to following R—knows what to do in order to follow R here. That is, we can in principle give an R-skeptic who accepts M/E instrumental reason to follow R, and if she can act on this instrumental reason—and Dreier assumes she can, since she accepts M/E—she will thereby follow R. Of course, getting an R-skeptic who accepts M/E to actually follow R on instrumental grounds is a contingent matter, dependent on getting her to desire to follow R and to believe that by doing φ she will follow R (96). But this introduces no principled obstacle to answering R-skeptics, only the possibility of contingent psychological obstacles. Dreier concludes that M/E has a "special status" among practical rules (95):

> Once you have (accept) the means/ends rule, what you need to get you to acceptance of other rules is one or another desire. But no desire will get you to the means/ends rule itself. (94)

I think the alleged asymmetry between M/E and other rules is in fact highly problematic. If we can offer an R-skeptic justifications to follow R that do not depend for their efficacy on one's already accepting R, then why think we could not also offer an M/E-skeptic justifications of M/E that do not depend for their efficacy on one's already accepting M/E? In particular, why think that the only method of getting the M/E-skeptic to follow M/E would have to be getting her to desire to follow it, a desire that Dreier alleges can only motivate by engaging one's acceptance of M/E itself? (More on this below.)

But the primary problem with the argument is that even if M/E does uniquely give rise to a regress of the sort Dreier describes, this fact does not appear to support what it was supposed to support. Dreier goes on to claim, seemingly on the strength of his regress argument, that someone "who doesn't accept the M/E principle cannot be given reasons of any sort"; and that whereas "there are (possible?) beings who can recognize reasons, who act on reasons, but are not moved by moral considerations," there are no possible beings who can recognize and act on reasons but are not moved by instrumental considerations (98). This sounds like the constitutivist instrumentalist's premise. The regress is supposed to show that acceptance of M/E is a condition of agency, whereas acceptance of other rules, such as moral rules, is optional so far as agency is concerned. (Talk of "rules" is to be taken loosely: Dreier means not to beg questions against e.g. virtue ethicists who ground ethical reasons in facts about good character but deny that having good character is a matter of following "general rules" codifiable in formulae analogous to M/E. See Dreier 2001: 27, n.2.) The problem is in seeing how the regress can show anything of the sort. The regress considerations, even if correct, appear in the first instance to show that one who rejects M/E is *stuck rejecting it*. The M/E-skeptic's requests for further and further justifications of M/E come to seem "empty" because there are, so the regress alleges, no possible practical justifications *of M/E* that the M/E-skeptic could heed. But this does not entail that an M/E-skeptic cannot accept other putative practical rules R even as she rejects M/E, and act on the non-instrumental reasons that such acceptance makes her receptive to. Acceptance of R may not be enough to make one receptive to reasons that could get one to accept M/E; but this does not entail that accepting M/E is a condition of acting for reasons, just as such.

For instance, the M/E-skeptic might accept the following rule "normative judgment internalism," NJI:

> **NJI** [For all φ] If one judges that, all things considered, one ought to φ, then has reason to φ.[64]

[64] I will continue to formulate practical rules as claims about reasons, instead of in imperatival form: the translation between the two is easy to effect, but the claim form allows me to talk about the rules, or their claims, as true. For a

A follower of NJI φ-s if her best judgment is that she ought to φ in the circumstances. To her, it is irrelevant whether φ-ing is a means to some end ψ she antecedently desires.

It does not matter here whether NJI is true, or even plausible. We are not supposed to assume that a given rule really is or is not authoritative: for the constitutivist, investigating the conditions of agency is supposed to be a way of investigating which rules are authoritative. All that matters is that the M/E-skeptic can follow NJI. And it is thus far unclear why she cannot. Since the regress concerns whether M/E-skeptics can be got to follow M/E, it does not even appear to address the question whether M/E-skeptics can follow other rules. Only if we presuppose that someone who rejects M/E cannot accept other rules R, and so cannot act for R-based reasons, does the regress even look to entail that, since the M/E-skeptic cannot be brought to act for instrumental reasons, she cannot act for any reasons at all. But to presuppose this would be to presuppose exactly what is at issue: that acceptance of M/E is a condition of agency.[65]

Dreier might admit this but hold that the regress at least shows M/E to be somehow special, since one who accepts M/E can be got to follow other rules R, but not vice versa. Perhaps such versatility in what a rule can get one to do, given some suitable psychological contingency, is even essential to agency; in this way, we could after all try mounting an argument for the claim that acceptance of M/E is a condition of agency. But this suggestion fails. For the regress cannot even support the alleged asymmetry between M/E and other rules without relying on the disputed claim that acceptance of M/E is a condition of agency.

Here is why. Let us continue to follow Dreier's model concerning the relation between one's "acceptance" of rules of reasoning and the sorts of reasons one is susceptible to: suppose that an agent's

---

recent defense of a slightly different version of NJI, see e.g. Ralph Wedgwood 2007, ch.1. For Wedgwood, making normative judgments of the form "I ought to φ" both (a) necessarily psychologically disposes one to (intend to) act as one judges one ought (insofar as one is rational), and (b) rationally requires one to conform one's intentions (or intentional actions) to those judgments. For criticism of NJI see e.g. Arpaly 2003, ch.2; Schroeter 2005.

[65] It is not in any case outlandish to think that agents might act through acceptance of NJI. When we wonder how *akrasia* is possible, we seem to implicitly assume that something like acceptance of NJI ordinarily moves agents.

receptivity to a type of non-instrumental reason is capturable through something like the thought that she accepts some principle of reasoning distinct from M/E. Then an M/E-skeptic who accepts NJI can be brought to "follow" M/E in just the same sense that Dreier claimed R-skeptics could be brought to follow R: through exploiting her acceptance of a rule distinct from the rule she rejects. Specifically, if the M/E-skeptic who accepts NJI can be brought to judge that what she ought to do in the present situation, all things considered, is to follow M/E; and if she also knows what M/E recommends in the situation (that is, if she knows that to follow M/E here she must φ, given that she desires to ψ and believes that by φ-ing she would ψ); then her adherence to NJI can also move her to follow M/E—though for NJI-based reasons. Just as an R-skeptic was brought to "follow" R, but to do so for instrumental reasons based on the newly instilled desire to follow R (where such reasons get a grip on the R-skeptic thanks to her acceptance of M/E), likewise an M/E-skeptic who accepts NJI can be brought to "follow" M/E, but to do so because she judges that she ought to do so now (that is, for a reason that gets a grip on her because she accepts NJI).

If this is right, there is no relevant asymmetry between M/E and other rules. Without presupposing that acceptance of M/E is a condition of agency, M/E-skepticism no more gives rise to a regress than does NJI-skepticism, since skeptics about one rule can always in principle be given reason to follow it in terms of the other. Indeed, we can always construct yet further rules with a suitable structure—a bootstrapping structure yielding justifications for anything provided a suitable psychological contingency—that an agent might conceivably follow, in terms of which we can justify other rules to her, including M/E. If so, the regress strategy is rendered useless.

We might object that this argument against Dreier's asymmetry claim makes a dubious assumption about what would suffice for answering the M/E-skeptic. Recall that for Dreier, one "accepts" a rule R when one is disposed to actually reason in accord with it, and to thereby act on the putative reasons that R picks out as applying to one (89). "Accepting" a rule thus implies being disposed to follow that rule. But not all senses in which one might be said to "follow" a rule imply that one "accepts" that rule. When one "accepts" M/E in Dreier's sense, one is disposed to do, *for instrumental reasons*, what M/E requires of one: one makes a consideration *p* one's reason to φ because of the way that *p* presents φ-

ing as a means to some end ψ one antecedently desires. In contrast, if one can be brought to "follow" M/E through one's acceptance of some other rule R, this can at most involve being brought to *externally comply with M/E* for R-based reasons—to do what one knows M/E to recommend, given one's beliefs and desires, but to do this for R-based reasons. We may call external compliance with M/E one kind of "following" M/E. But it is not "acceptance" of M/E. Thus we cannot after all answer the M/E-skeptic in the sense requisite for getting her out of the regress: we cannot get her to "accept" M/E through exploiting the fact that she accepts R.

Though sound, this last line of thought does not help Dreier salvage the asymmetry he needs. The difference between kinds of "following" a rule—the kind involved in "acceptance" on the one hand, and the kind involved in external compliance on the other—is a perfectly general feature of practical rule-following. By the same reasoning, neither can R-skeptics be brought to accept R through their acceptance of M/E. The most that acceptance of M/E can achieve is to get one to do, for instrumental reasons, what one knows R to require of one. For instance, one might be brought to do what one knows the rules of justice to require of one, but to do this for the instrumental reason that one desires a good reputation and believes that by violating the rules of justice one is likely to acquire a bad reputation.[66] Not that one cannot come to accept new rules in any way; but it is hard to see, as a general matter, how prior acceptance of some rule, together with proffered practical justifications of the sort to which one is already receptive thanks to that prior acceptance, could on its own achieve a transition to acceptance of new practical rules. To "accept" a rule in Dreier's sense is to have a disposition of practical thought, a disposition to reason in a certain way and to take certain things as reasons. It looks to be a general feature of practical rule-following that one cannot practically reason one's way into "acceptance" of a new rule. If so, Dreier's regress arises no less with R than with M/E.[67]

---

[66] Cf. Glaucon's suggestion at the beginning of *Republic* II.

[67] Perhaps a transition into acceptance of a new rule could instead be a "conversion," or a result of habituation and training of one's dispositions of practical thought over time (McDowell 1995a).

In sum, it is thus far unclear how Dreier's regress considerations could help support the claim that acceptance of M/E is a condition of agency. Reflection on the regress thus far only supports the general conclusions that (a) in practical deliberation, one must follow some rule or other; and (b) to "follow" a rule in the sense of reasoning in accord with it and acting on the putative reasons it picks out as applying to one, it is not enough to have some intentional attitude towards following the rule in question—e.g. a desire or intention to follow it, or a belief that one should—that then engages with some *other* rule one is following. This would be mere "external compliance," not reasoning in accord with the rule. But these general conclusions about practical rule-following do not favor M/E over other rules.

We may make a further observation about practical rule-following as such. Not only could reasoning in accord with a rule R not consist in having an attitude towards R *that then engages with some other rule one is following*; but furthermore, neither could reasoning in accord with R consist, simply, in having an attitude towards following R. Call the view that reasoning in accord with R simply consists in desiring or intending to follow R, or in believing that one should, the "Intentional View" of rule-following.[68] And now suppose, for example, that I desire to follow M/E. M/E is general: it tells me what to do in a range of situations. So acting on the desire to follow M/E is not something that I can just straightaway do. Instead, it seems that I must somehow recognize, whether explicitly or implicitly, that I am in the sort of situation to which M/E applies. In M/E's case, this means that I must recognize that I have the end-desires and means-ends beliefs that give M/E application. But I must also somehow put together this recognition with my desire to follow M/E, so that I then do what M/E tells me to do in the situation. And this seems to involve an inferential step. Now, Dreier supposes that the relevant inference would itself have to be an instance of M/E-reasoning. In fact it seems not to be, since my recognition that I have the means-end beliefs and end-desires that give M/E application is not itself a means-end belief. But *whatever* my rule of reasoning is in getting me from my desire to follow M/E to actually conforming to it, the present point is that we could not in turn account for *this* instance of reasoning in terms of a

[68] The term 'Intentional View' is from Boghossian 2008a. The argument that follows is a paraphrase of his argument against the Intentional View (127-8).

desire to follow its rule. Such a move would start a regress: in order to act on the new desire to follow the relevant rule, one would again have to take some inferential step or other, which on the Intentional View would require a further desire (or intention etc.) to follow a rule, and so on *ad infinitum*. So on the Intentional View of rule-following, one could never actually follow any rules. The Intentional View must be false.

I think something like this thought is the truth in the area of Dreier's regress argument. But importantly, it does not support the conclusion that M/E is a condition of agency. It only supports the conclusion that the notion of desiring (or intending etc.) to follow a rule, regardless of the rule in question, cannot furnish an account of practical rule-following as such. So we need another account. This general conclusion is compatible with the thought that, if we had a better, alternative account of practical rule-following, we could apply that account directly to rules distinct from M/E, so as to make sense of how we can reason in accord with them, even as we reject M/E.[69]

The foregoing helps to clarify how any argument in the area of Dreier's must be construed if it is to support constitutivist instrumentalism. The argument must try to directly establish the strong claim that acceptance of M/E is a condition of agency:

Practical Paralysis Someone who does not accept M/E cannot act for reasons at all, and cannot reason practically. (Not so for other rules.)[70]

In terms of the foregoing, this means showing that someone who rejects M/E cannot accept other rules R, and act through this acceptance. And to establish this, our attention must shift entirely away from the

[69] That is, we should not assume that there must be some foundational rule through which we follow all other rules. (That M/E is such a foundational rule is in effect the view that Dreier is trying to argue for.) The very same account that makes sense of reasoning in accord with the alleged foundational rule could instead be applied to other rules directly. A thought along these lines also destroys Korsgaard's 2009 rationale for the claim that we need a "hypothetical imperative" because we need a "general principle of practical application" (47, 65-67). §2.4 below says more against the idea of foundational rules.

[70] Since my focus is on arguing that acceptance of M/E is not a condition of agency, I focus only on whether it is, not on whether acceptance of some other practical rules might (also) be such a condition.

question *how to justify M/E to an M/E-skeptic*, and onto the question *whether something in the nature of practical rule-following as such entails Practical Paralysis*.

A natural avenue of argument for the constitutivist instrumentalist to pursue is to try to show that, even though the notion of desiring *to follow a rule* cannot furnish an account of practical rule-following as such, nonetheless desiring *some* end or other is necessary for practical rule-following, in some way such that Practical Paralysis follows. I consider this strategy below, showing that it fails. Instead, further reflections on practical rule-following support the possibility of accepting other rules R even as one rejects M/E: there can be non-instrumental agents.

2.3 IN WHAT SENSE MIGHT END-DESIRES BE NECESSARY FOR PRACTICAL RULE-FOLLOWING?

It might seem that in appealing to the possibility of NJI-following, I illicitly denied the "methodological axiom that whatever is missing when someone has a belief and lacks a certain motivation is a desire of some sort or other" (93). For one who accepts NJI looks to be moved by a belief, not desire—namely, the belief that, all things considered, one ought to φ. Perhaps, then, if we knew on independent grounds that in being moved to act one must be moved through some antecedent end-desire, we could also show that this condition of agency entails a further condition: that one must accept M/E.

The claim that one must always be moved by antecedent desires amounts to the claim that, while we may trivially ascribe a desire to φ to any agent as a logical consequence of her φ-ing, this trivial ascription does not sufficiently respect the "methodological axiom": instead, one's φ-ing must be

motivated by some further desire to ψ.[71] But (A) why believe the "methodological axiom," construed in such a strong way? And furthermore, (B) exactly how might it help the constitutivist instrumentalist?

Let us address (B) first. It is clear that alleging that practical rule-following requires that agents have desires, or more carefully, the contents or expressions of desires, as "premises" in their reasoning will not help the constitutivist instrumentalist. For even if desires were required as premises, this would not entail that one must accept M/E in particular. Consider:

> **Mother Says** [For all ψ and all φ] Whenever one desires anything, ψ, one should do whatever, φ, one believes one's mother would want one to do in this situation.

Someone who accepts Mother Says and desires something, ψ, could be got to φ by getting her to believe that her mother would want her to φ in this situation. Again, it does not matter whether Mother Says is true or even plausible. It might be a silly rule. Some agents are silly. That is not yet to show that they are not agents.

So if practical rule-following depends on desire in some way that entails Practical Paralysis, this dependence on desire must take some form other than dependence on desires (or their expressions) as "premises."[72] What, then, is the nature of this dependence? We already saw that reasoning in accord with a rule cannot just consist in a desire to follow it: the Intentional View of rule-following is false. There must be some other account. The only well-known alternative to the Intentional View is the "Disposition View" that one follows a rule through manifesting a disposition to act in the ways that the rule tells one to act, in situations to which the rule applies.[73] The general idea of such a disposition roughly matches Dreier's notion of "accepting" a rule. But stated in such vague terms, it is clear that the Disposition View does not support Practical Paralysis. Everything depends on *what it is* to be disposed to follow a rule in

---

[71] On "trivial" ascription of desires, see Nagel 1970: V 1-2.
[72] In any case, why think that, to act even through acceptance of M/E, one would *need* to have actual desires, or their expressions, as premises? It seems that one who accepts M/E might be moved, through that acceptance, to act on a belief that she desires to ψ, even if that belief is false. See Schueler 2009.
[73] Cf. Boghossian 2008a: 130.

the relevant sense; and thus far it seems that one could be disposed to follow other rules even as one rejects M/E.

However, Michael Smith's influential "Humean Theory of Motivation" (1987; 1994, ch.4) helps us to formulate a suggestion for how to flesh out the Disposition View so that it does support Practical Paralysis. According to this *Sophisticated Humean* view, as I'll call it, practical reasoning depends on end-desires, themselves dispositionally understood, in a way that shows that practical reasoning must be reasoning in accord with M/E. Smith's argument for the "Humean" theory of motivation is also an argument for the strong construal of the "methodological axiom." So in addressing it we will also address question (A), above. Let us turn to this possibility.

The "Humean" theory of motivation consists in the following thesis:

> **(P1)** In φ-ing, one's motivating reason for φ-ing must consist of a desire for an end ψ together with a belief that by φ-ing one will ψ.[74]

Smith's argument for (P1) is that it is the only view of "motivating reasons"—the reasons that move one, unless defeated by other motivating reasons (1987: 38)—that can make sense of motivation as "the pursuit of a goal" (44-5 ff). The argument involves three central claims. The first is that motivating reasons "have the potential to explain [one's] behavior," and are therefore psychological states of the agent (38). The second is that motivation is indeed "the pursuit of a goal," and that hence the psychological states in which one's motivating reasons consist must themselves be "constituted by goals" (or by the having of goals) (38, 44-5). The third claim is that (a) desire is a functional state "disposing the subject [who desires that *p*] to bring it about that *p*" when in conditions C, where C includes beliefs about how to bring it about that *p* as well as a perception that, presently, not-*p*; and that (b) given this "functional role" or "dispositional" account of desire, desiring an end just is having a goal: to desire that *p* is, *ceteris paribus*, and *inter alia*, to try to bring it about that *p*, and this just means having as one's goal the state of the world in which '*p*' is true (54-5). Since being moved by motivating reasons is being

[74] I have simplified Smith's formulation somewhat; see 1987: 36. But nothing turns on this.

moved by (the having of) a goal, and since having a goal is having an end-desire, it follows that end-desires must be constituents of motivating reasons. The other constituents of motivating reasons are means-ends beliefs, together with a perception that one's end does not yet obtain.

Now, it is important that in speaking of "motivating reasons," Smith does *not* mean one's "premises" in reasoning. An agent's motivating reasons are not the contents of her deliberations, the considerations on which she acts: they need not figure in the "foreground" of one's deliberations at all. Rather, they figure in the "background," where their functional role is to explain one's choice in view of whatever explicit justifications one adduces for that choice (Pettit & Smith 1990: 568-9), therein explaining the course that one's reasoning takes in view of one's premises. Relatedly, though, one's premises may, in turn, reflect one's motivating reasons in some way. One might only think that ψ-ing would be fun or desirable, and act on the basis of such considerations, if one background-desires to ψ; and one might only be likely to adopt, say, "φ-ing is my duty now" as a premise in one's reasoning if, and because, one background-desires to do one's duty (Smith 1987: 53-4; Pettit & Smith 1990: 576-7). In this way, the considerations on which one acts—one's foreground justifications for a given course of action—can be various in content and in phenomenology: one can appear to be, and see oneself as being, responsive to various values, duties, and so on. But what ultimately explains one's seeing such considerations as supporting a given course of action is that they reveal, perhaps implicitly, that course of action as satisfying a background end-desire.

Given this "background" understanding of (P1), (P1) can be construed as offering an account of the very idea of a disposition of practical reasoning. For the schematic idea of such a disposition is precisely the idea of something in the "background" of one's practical reasoning, distinct from one's premises, and explaining, first, why considerations *xyz* figure as premises in one's reasoning at all, and second, one's course of reasoning in their light. Moreover, the account looks to entail that unless one accepts M/E, one could not deliberate at all. For acceptance of M/E looks to be precisely the "functional role" that Smith assigns to background end-desires when combined with (background) means-ends beliefs. Smith's agent need not have any thoughts *about* M/E itself; rather, the idea that one's actions

must be chosen as ways to satisfy a background end-desire, regardless of the content of one's foreground justifications for action, can be understood as just one way of saying that the basic *form* of one's reasoning, and of the practical justifications that figure as one's premises, must be instrumental. Insofar as one acts on the basis of practical justifications *xyz*, one acts on the basis of considerations that move one only *qua* steps towards zeroing in on φ-ing as a means to one's background end, and so, on considerations that move one only *qua* instrumental justifications. We might say that whenever one is moved to φ on the basis of practical justifications *xyz*, one is moved by an implicit conception of φ-ing as a means to some ψ one background-desires.[75]

The central question in assessing this Sophisticated Humeanism is why we should accept that the "background" psychological states that move one, whenever one reasons practically, must be end-desires and means-ends beliefs embodying acceptance of M/E. Even if we accept Smith's characterization of the specific "functional role" of background end-desires, so that such desires can only move one through embodying acceptance of M/E, why think that what is in the "background" of one's reasoning must always be some such end-desire?

Smith explains that motivation as such just is the pursuit of a goal, and that to make sense of this, we must construe "motivating reasons"—one's background psychological states—as themselves partially constituted by (the having of) goals. This was the second claim in Smith's argument for (P1), now interpreted in terms of the "background" understanding of the role of "motivating reasons." But the claim is question-begging in favor of the instrumental picture of reasoning. It may be true that, in being moved to act, we are moved *to* pursue some goal: we are moved to do something or to pursue something. If I decide to φ, then we may say that φ-ing is now one of my goals, which I am pursuing when I am doing φ.

---

[75] I invoke the notion of an implicit conception here only as a way of gesturing towards a picture of practical rule-following as a phenomenon in which an agent somehow grasps a way of going on, and goes on in that way, seeing a certain kind of significance in specific "premises," without her having to explicitly represent, in conscious awareness, either her way of going on or the significance that her going on in this way shows her to attach to her "premises." I do *not* mean to imply that having an implicit conception of a conclusion as following according to some rule P constitutes knowledge of the validity or goodness of P (cf. Peacocke 1998). More on this picture of practical rule-following as involving implicit conceptions below.

But this does not yet entail that what we are moved *by*, in being moved to pursue some goal, is some further (background) goal. To suppose that it does is to presuppose Sophisticated Humeanism: it is to presuppose that one must always choose φ-ing because one has an implicit conception of φ-ing as instrumental to some further end. We need some independent reason for believing the conditional claim that, if motivation is "the pursuit of a goal" (1987: 44), then "motivating reasons" must themselves be constituted by goals. But Smith himself gives no such reason. He merely takes it to be a "conceptual truth" about motivation (1987: 55).[76]

Stephen Finlay (2007) does propose a strategy that looks like it might help establish the requisite conditional claim.[77] The strategy proceeds by appeal to the notion of *voluntary* behavior:

> [I maintain that] all voluntary behaviour is caused non-deviantly by desire (i.e. it is motivated). Given my account of desire, this is just to say that all voluntary behaviour is teleologically caused, the intentional result of mentally aiming at some end. But why should anyone accept this […]? I submit that intuitively this just is the essential difference between the behaviour, both physical and mental, that comes upon us (or that we do without volition), and that which we actively and voluntarily perform. If some behavior B is not an intentional result of my aiming at it (or of my trying to produce it)—directing my motions and thoughts in the ways I think will or may lead to B—then I cannot recognize it as something that I do voluntarily. (2007: 231)

Finlay considers the possibility that "inferential processes" might be counterexamples to his thesis that the voluntariness of X depends upon X's being an "intentional result" of my "aiming at some end." Theoretical reasoning does not seem to rely on desire, yet it seems to be a paradigm example of something we actively do, not something that merely "comes upon us" such as a "reflex to kick when the knee is tapped" (231). And this may suggest that there are "similarly *practical* inferences, with actions as conclusions, that do not depend on desires" but are nonetheless voluntary (232).[78] Finlay responds, however, that even theoretical reasoning does depend upon a prior desire for its voluntariness. This is his argument:

> [We must] distinguish evidence or 'reasons for belief' from reasons to *form* beliefs. A justified belief always constitutes evidence for its logical consequences but does not always provide a

---

[76] Cf. 38, 44-5, 55, and 58-9: Smith simply asserts the alleged conceptual truth, responding even to Nagel's 1970 objection to it by re-asserting it.
[77] I thank an anonymous referee for pressing me to address Finlay's work.
[78] Finlay attributes the objection from reasoning to his notion of voluntariness to Wallace 1990 and Smith 1987-8, as well as to personal communication with Smith.

> reason (i.e. normative pressure) to form such beliefs, simply because many of the logical consequences of our beliefs are utterly trivial. […] So the existence of normative reasons to form certain beliefs is conditional on more than simply logical or evidentiary relations. My perceiving myself to have such a reason, and my being motivated to form the belief, I argue, depends upon my intellection being motivated by some desire such as the desire to know about subject X, or (more accurately) to settle whether something is the case. (2007: 232)

Although Finlay does not explicitly extend the argument to practical reasoning and acting for reasons, it is clearly intended: he holds that any behavior, including practical reasoning, "can only be voluntary if it is motivated by some further desire" (234-5). If Finlay were right, then being moved *to* pursue a goal would indeed have to involve being moved *by* an antecedent desire for a goal, in order to be voluntary. But Finlay's argument is puzzling. It seems true that facts about logical implication do not straightforwardly constitute norms of reasoning: I am not yet violating a norm of reasoning if I do not move from my current beliefs to all of their logical implications, forming e.g. infinitary disjunctive beliefs.[79] But it does not follow that my voluntarily following norms of reasoning depends on some prior desire I have. It may be true that something beyond logic is needed for epistemic obligation; but this claim says nothing about desires, nor does it say anything about what it is to voluntarily follow epistemic norms. Yet beyond the argument quoted, Finlay says nothing to support his intuition that voluntariness depends upon prior desire.[80]

Finlay is surely right to hold that reasoning is "voluntary" in some sense contrasting with automatic reflexes and processes that merely happen to us or "come upon us." And there is an element of truth in Finlay's definition of the voluntary: it does seem that, whenever I do B voluntarily, I must in some sense aim to be doing B. But this does not entail that voluntary behaviors must be directed towards some goal *beyond themselves*. That I must "aim" to be doing B in order to be voluntarily doing B may only mean that, in voluntarily doing B, I cannot think that I am doing something else, nor be completely ignorant that I am doing B. In order to be voluntarily doing B, I have to *mean to be doing B*. We might put the point by saying that voluntarily doing B is essentially a *self-conscious* activity: I would not be

[79] Cf. Harman 1999: 2.1.1.

[80] He does respond to an "objection" that flatly denies his thesis, but only by "conceding" that "automatic" inferences that merely "strike" you don't require prior desires, and "maintaining" that, nonetheless, voluntary inferences that you actively "draw" do (2007: 232). If there is an argument here, I do not understand it.

voluntarily doing B unless I was at least implicitly aware of myself as doing B, and in this sense aimed to be doing B.[81] Importantly, it seems that in order to voluntarily do B, it must be in some sense *through* my implicit awareness of myself as doing B that I do B. My awareness is not just an accompaniment to my doing B, where B itself springs from some possibly "alien" cause, involuntarily; rather, my awareness is essential to my doing B at all.[82] This gives a sense to the claim that, to be voluntarily doing B, my "aiming" to be doing B must "direct" my doing it. None of this entails, however, that in self-consciously reasoning I must also be aiming at some *further* goal. The mark of the voluntary may be self-consciousness, not directedness at an end beyond the voluntary activity itself.

This of course falls short of an argument that this is how we *must* think of "voluntary" behavior. But the outlined proposal at least improves on Finlay's in explaining why, if I am actually doing B voluntarily, this fact cannot be completely alien to me: I must be somehow aware of it. The explanation is simply that my knowledge of my doing B is partly constitutive of my voluntarily doing B. Nothing in the notion of being "non-deviantly" caused by prior desire provides a competing explanation.

If this is right, then the notion of the voluntary does not provide independent reason to believe that, to be moved *to* pursue something, φ, one must be moved *by* some antecedent desire for a goal, ψ. But someone might object that construing of the voluntary in terms of self-consciousness is too intellectualized, and so not a viable alternative to Finlay's conception. For surely, the objection goes, one's reasoning in accord with a rule R can be voluntary in the sense of being distinct from something that merely happens to one, without one's knowing *that one is reasoning in accord with R*, under that description. As Wittgenstein says, we need no explicit formulae to be able to cotton on to the rule in, say,

---

[81] This is why the sincere answer "I wasn't aware I was doing that," in response to the question "why are you doing B?", reveals one not to have been acting intentionally or, in Finlay's vocabulary, voluntarily. This aspect of intentional action is a central theme in Anscombe's 1957 *Intention.* With reasoning, however, the "why"-question may deceive: one can voluntarily reason in accord with a rule R without knowing, *under the description "reasoning in accord with R,"* that one is doing so. (See below.)

[82] This distinguishes the sort of self-consciousness at issue in "voluntary" reasoning and action from a receptive awareness one may have of, say, one's own emotional states. Cf. Anscombe's thought, which she attributes to Aquinas, that the knowledge one has of one's own intentional actions is "the cause of what it understands" (1957: §48).

a number series and to gradually start to follow it (*Philosophical Investigations (PI)* §§143-152). Nor would it help, on its own, if a formula did occur to me: it might be gibberish to me, and I might be unable to follow the rule it states (*PI* §152). It is a virtue of the Disposition View of rule-following that, at least in the abstract, it seems not to require one to employ explicit formulations of the rules one is following, as one is following them. We should beware fleshing out the Disposition View in terms that sacrifice this virtue.

Nothing in the appeal to self-consciousness, however, conflicts with Wittgenstein's insight. In fact, the kind of awareness one has of oneself as reasoning in a certain way, as one is reasoning in that way, is best understood in terms of Wittgenstein's insight. It is true that, in cottoning on to a rule R and coming to follow it, I need not acquire a conception of myself as reasoning in accord with R, under that description. Rather, what I acquire, and then exercise in following a rule, is *knowledge how to go on* by the lights of the rule: I come to understand how to go on, and my ability to actually follow the rule is in part constituted by this understanding (cf. *PI* §§147-154ff.). This know-how does not require me to be able to articulate an exact formula for how I go on, nor is its exercise, therefore, well conceived as a matter of my mechanically deriving consequences from a formula reflectively available to me. Instead, it is enough that I have an implicit conception of certain conclusions as called for in my situation by the lights of the rule I am following—a conception that (i) guides my actual conduct when I follow the rule, tending to conform my conclusions to the rule's dictates; and that (ii) I can try to articulate by going over what I have done, or by giving further examples of what I would conclude in similar instances and why.

This is the sense in which it cannot be completely alien to me that, and how, I am reasoning, as I am reasoning. This is *somewhat* "intellectualized," to be sure, in being somewhat cognitively demanding. Condition (i) includes the requirement that, for one to be "following" a rule, one's conduct must approximate *correctness* by the rule's lights. This means that one's implicit conception how to go on must not be just *any* conception, but a broadly correct one: the central notion is *know*-how, not just belief-how. But there is good reason for this requirement. Someone who is guided by *a* conception how to (say) add, but a completely wrong one, could not really be said to be following the rule of addition at all. (Perhaps it

would not even count as a conception how to add.) There is likewise good reason for condition (ii). If I cannot in *any* way explain how I get to my conclusions, then it seems that we would be justified in doubting whether I am really reasoning at all. The point is that nothing in these conditions entails the objectionable claim that, in reasoning in accord with R, I must also know that I am reasoning in accord with R, under that description.

So we can reject Sophisticated Humeanism while avoiding fleshing out the Disposition View in objectionably intellectualized terms. In being disposed to follow a rule, I am disposed to exercise a specific ability, an ability to go on broadly correctly by the lights of a rule, in the circumstances to which the rule applies (Cf. *PI* §150).[83] It is the concept of such a disposition that an account of "accepting" a practical rule must explicate further. And nothing in the idea of such a disposition thus far suggests that a specifically instrumental norm is implicit in it. An implicit conception of φ-ing as a means to some background end ψ is just one kind of implicit conception one might have guiding one's reasoning towards φ-ing. Another is an implicit conception of φ-ing as, say, what one's mother would want one to do in the situation.

I have been arguing that neither the fact that action and reasoning is voluntary, nor the fact that in practical reasoning we are moved *to* pursue some goal or other, implies Sophisticated Humeanism. If end-desires are necessary for practical reasoning in some way that entails that acceptance of M/E is a condition of agency, we have yet to see how. But here is a final suggestion: perhaps practical reasoning must at least be minimally end-governed in the sense that it is governed by the end of "deciding what to do." Candace Vogler suggests this, claiming that without such an end, there is "nothing to engage or direct practical reason" (2002: 166-170). In response, however, even if there is a sense in which practical reasoning must always be aimed at deciding what to do, this fact cannot on its own support an instrumental norm of reasoning, or indeed any other norm. This is because *any* decision would satisfy the

[83] Can I possess the ability without the disposition? I think this is a tough and interesting question, and its resolution ultimately impacts the fate of the Disposition View. It also has implications for how to understand freedom in reasoning. But this topic is beyond our scope here.

end of "deciding what to do": no decision could count as going wrong in relation to that end. In order for a putative practical norm to be genuinely normative, it must at least be possible for there to be accord and discord with its dictates, so that not just anything counts as following it. (Though there is also *more* than this to normative authority: genuinely authoritative practical rules not only claim something contentful about how we should reason and decide, but also, claim something *true* about how we should reason and decide.) The crucial question that norms of practical reasoning are supposed to answer is *how* one should go about deciding what to do, and (so) which decisions and actions are called for given one's situation. To be a putative norm of reasoning, then, rather than merely a description of what happens whenever one reasons, a rule should have something to say about that question (in the contexts to which the rule applies).[84]

If the foregoing is right, then a Sophisticated Humean analysis of practical rule-following is not compulsory: nothing in the nature of practical rule-following as such requires it. We may say, perhaps, that any disposition to follow a rule is "desire-like" in the sense of helping to move one to act (cf. Dreier 1997: 93-4). But the substantive question is how to account for such dispositions; and in pursuing an account, we have thus far encountered nothing that supports Sophisticated Humeanism. We have found no essential explanatory role for background end-desires to play in accounting for practical rule-following, and hence no reason to think that the only rule one might follow in practical reasoning is M/E. Instead, the central notion of practical rule-following appears to be that of exercising knowledge how to go on by the lights of whatever rule one is following—whether the rule is M/E, NJI, Mother Says, or something else.

However, this result may leave us unsatisfied. For neither have we thus far uncovered anything in the nature of practical rule-following that positively speaks against the Sophisticated Humean's account

---

[84] Compare Finlay's suggestion that theoretical reasoning is aimed at the end of "settling whether" *p* (2007: 232, quoted above). Whether one settles on *p* or ~*p*, one's answer would satisfy this end. This suggests that the operative notion of end-directed reasoning is normatively empty: even if reasoning itself *were* end-directed in Finlay's sense, this fact could not support norms of reasoning. Cf. Finlay 2007: 236-7; 2006. On the idea that norms must be violable to be norms, see e.g. Korsgaard 1997: 228 and Dreier 1997: 91.

of it. In particular, we have thus far given no reason why the Sophisticated Humean cannot in turn account for the sort of know-how and its exercise involved in practical rule-following in her terms. If we could find apparent cases of such know-how and its exercise that the Sophisticated Humean cannot account for in her terms—compelling appearances that Sophisticated Humeanism cannot save—then we would have to conclude that Sophisticated Humeanism is false as a general account of practical rule-following.[85] This is our task in the final section. The argument will also show that there are no other reasons, besides Sophisticated Humeanism, to accept Practical Paralysis. There are compelling cases of non-instrumental agency—cases that, further, tentatively support content skepticism about constitutivism.

2.4 CHARACTER DISPOSITIONS AS SOURCES OF ACTION: THE IMMEDIACY OF PRACTICAL RULE-FOLLOWING

To start with appearances: take the friendly person, who does what she does "from" friendship. We need not suppose that such people are common, nor even that any such person actually exists. We merely need to suppose that she is possible, as she seems to be, and ask: what account of action and deliberation best makes sense of her, as well as of other analogous appearances of acting "from" a character disposition?[86]

Suppose our friendly agent's friend is sick, and so she decides to visit him, doing so from friendship. Acting "from" character traits such as friendship seems to be a paradigm case of having and exercising know-how in practical reasoning. The friendly person knows how to go on by the lights of friendship and exercises this knowledge. In doing so, we may suppose that her thinking about what to do

[85] I mean this procedure to be reminiscent of Kant's in the *Groundwork*: I think we grasp the relevant appearances in part through common ethical knowledge, and the philosophical task is to show how they are possible without distortion. But instead of finding a common principle as the principle of willing for all agents, I think we can repeat the epistemic procedure for many different putative principles of "willing," without admitting that any of those principles therefore belong to the metaphysics of agency as such.

[86] The idea that character dispositions might be sources of action is of course familiar. What is not sufficiently appreciated is how this idea speaks against the idea that agency as such involves reasoning that is instrumental in form.

is informed by her "background" disposition of friendship, a disposition that directs her choices in the light of her "premises," as well as directs her to make certain considerations her "premises" in the first place. (The content of the premises might be something like "He needs help" or "He needs cheering up"; on this, all sides can agree.) The Sophisticated Humean would analyze the relevant background disposition as the disposition to follow M/E, embodied in some background end-desire(s) or other, such as a desire for (this) friendship, for being a good friend, or for cheering up one's friend; together with (background) means-ends beliefs.[87] But can these Sophisticated Humean resources explain what it is to act "from" friendship—to know how to go on by the lights of friendship and to exercise this knowledge? Is the form of the friendly person's reasons for action aptly thought of as instrumental?

Let us start by considering the helpfulness of the notion of a background end in this context. To act "from" friendship, one must *have* friendship as a character disposition. But clearly one can act in instrumental pursuit of any proposed background goal without yet having, and so without yet acting "from," friendship. An obvious such case is when the specified background goals are adopted in the service of yet further goals: one may seek to cheer up a friend in order to look good in one's own estimation, or to pursue (this) friendship in order to forge business connections. In such cases, the character structure revealed by one's choice, what drives one's choice, is not friendship but something else, such as petty self-love or ruthlessness. A different case is when one pursues e.g. (this) friendship, or being a good friend, for its own sake and not for the sake of some further goal. But again, someone might adulate friendship as the highest good, seeking it solely for its own sake in all one's actions, without yet necessarily having friendship as a disposition: it might be a distant and as-yet-unachieved goal.[88] Indeed,

---

[87] This is suggested by Pettit & Smith's remarks at 1990: 567-7. Something like it seems to have also been Anscombe's picture, in *Intention* §41, with regard to acting "from" duty. Though not usually thought of as a particularly "Humean" figure, Anscombe seems to have shared the rough "background end" picture of the "motor force" explaining the tenor of any given episode of practical reasoning. Cf. Anscombe 1995: 6-7.

[88] This argument is inspired by Stocker 1981. In formulating it, as well as with most everything else in this dissertation, I was much helped by comments from Kieran Setiya.

Smith's own analysis of desiring a goal suggests that the goal *must* be unachieved: for Smith, a desire that *p* is a disposition to bring it about that *p*, in the presence of a perception that, presently, not-*p*.[89]

Nor will it help to suggest that the friendly person seeks some other goal, besides friendship itself, solely for its own sake. Wanting to, say, cheer up one's friend for its own sake and not for something further may be characteristic of the friendly person, when doing so is appropriate by the lights of friendship. But someone else may wish to cheer up one's friend always and without condition, without heed to whether one's attempts to do so are becoming tiresome. Having friendship involves sensitivity to changing situations: the person with friendship may come to see reason to change her specific goals where the person who lacks friendship but shares the initial goals would be blind to those reasons. The Sophisticated Humean might respond that such changes in one's specific goals can be explained by one's noticing that the specific goals in question have stopped serving one's overall goal of friendship, or of being a good friend. But since one can have and pursue those "overall" goals without yet having and acting from friendship, it is hard to see how appeal to them could help at this stage. In general, it seems that the subjunctive conditionals that a correct character attribution supports—statements concerning what the person would do, or would tend to do, in some hypothetical situation—differ from those that any given set of goal-attributions supports on its own. Indeed, mere goal-attributions may not support many such subjunctives at all: what the honest but reckless person does when she aims to make a quick buck is quite different from what the dishonest but careful person would do given the same aim.

The Sophisticated Humean might object that she need not be stuck with the narrow notion of desiring an end that Smith assumes. Smith supposes that ends desired are always "finite": unachieved while one pursues them and expiring when one achieves them, such as getting the camera from upstairs. But there are also "infinite" ends, such as healthy living or honoring one's parents, which one may

[89] One might object that the friendly agent need only *perceive* her goal of friendship to be unachieved for that goal to continue to motivate her: the goal might be, in fact, achieved. But it would be an odd picture of friendship on which acting from it depended on having such a misperception.

continue to pursue even as one already achieves them at every moment of the pursuit.[90] According to this objection, friendship *could* be a matter of pursuing specific end-desires such as the goal of being a good friend, if that end is an infinite one: the end could be achieved even as it continues to guide one's pursuit of itself, and of subsidiary goals.

But this does not help. Even though one may continue to desire infinite ends even as one has already achieved them, nonetheless it is also possible to desire infinite ends *without* yet having achieved them. I can desire, say, friendship or healthy living without yet having actually achieved them. If only achieving them were as easy as desiring them! It is only when one has actually achieved the object of the desire for friendship that one has friendship and can act "from" friendship. And it is precisely what "having" friendship is, so that one might act "from" it, that we are trying to understand.

The discussion thus far suggests that character dispositions such as friendship are not capturable in the idea of having specific end-desires and pursuing them through one's acceptance of M/E. But our last way of putting the problem—that even if one aims at friendship as an infinite end, one does not yet have friendship, and so cannot act "from" it, unless one has actually achieved that aim—suggests that the central issue is one of know-how. Having the goal of friendship and pursuing it through instrumental reasoning is insufficient for having and acting from friendship, we might suggest, simply because it is insufficient for *knowing how to achieve that goal*. If this is the complaint, then the Sophisticated Humean ought to respond as follows: *Of course* it is possible to have and pursue goals such as friendship or being a good friend without yet having friendship; but this is simply because it is possible for one's instrumental reasoning to be either bad or based on false beliefs. If one reasoned instrumentally well, and one knew the proper means to being a good friend, then one would take those means, and there would be no problem: one would, in fact, be a good friend.[91]

---

[90] The examples are Sebastian Rödl's, in his 2010 "The Form of the Will": 146-148. In formulating the present objection, I am much indebted to Rödl's discussion, and to Stephen Engstrom, who pressed me to address the topic.
[91] Reflection on some comments from Karl Schafer helped me to formulate this objection, though I am not sure if it is precisely the objection he would make.

But our problem remains. Consider again the possibility that one can pursue the infinite ends of friendship or health even while one has not yet achieved them. If it were not possible to do this, then one could not strive towards friendship or health, conceived of as infinite, non-expiring ends, while lacking friendship or health. But surely one can do so. In fact, surely it is possible for one to strive towards friendship or health *well*: it must be possible for one's means to get one there. One is not stuck forever being surly or obsequious, nor is the only way to stop being so to wait for a miraculous transformation. One can take steps to achieve these goals, and these steps might eventually get one there. The problem for the Sophisticated Humean is that such successful striving towards friendship or health seems to be an instance of good instrumental reasoning in pursuit of the infinite end of friendship, where one's reasoning involves true beliefs about the means. But as before, the one who strives towards friendship does not—or at any rate need not—yet have it. So the problem remains. The Sophisticated Humean's resources leave us short of a way to understand what it is to have and act "from" friendship, since they give us no way to distinguish between the one who merely strives towards friendship with eventual success, and the one who already has it.

It will not help to assert that what distinguishes the friendly person from the merely striving one is that the friendly person has already achieved her aim, even as she continues to pursue it. For this is just to assert that she has friendship. And what it is to have friendship as a practical disposition, so that one may act "from" it, is just what we have been trying to grasp, unsuccessfully, through the Sophisticated Humean's resources.

One might object that the foregoing ignores an important distinction between types of means: constitutive means, which are ways of *realizing* or *actualizing* an end (of e.g. being a good friend), and efficient means, which merely get one *towards* the end. What distinguishes the friendly person from the striving one is that the friendly person takes correct constitutive means to the end of friendship, whereas the merely striving one takes efficient means. In response, while it may be true that the friendly person takes constitutive means to being a good friend in the sense that, in acting "from" friendship, she acts in a way that helps to continually constitute her as friendly—her actions are friendly actions—nonetheless the

striving person can also take constitutive means to friendship: by acting as the friendly person would, and in other ways emulating her, she can gradually constitute herself as friendly, and in this way actualize friendship in herself.[92] (Compare Aristotle's remarks on moral education in Book II of the *Nicomachean Ethics*.) The point is that the striving person is not yet there, whereas the friendly person who acts "from" friendship is.

We may put the point in terms of the know-how of the friendly person. In knowing how to go on by the lights of friendship and exercising that knowledge, the friendly person's know-how eludes capture in the idea of knowing how to achieve the goal of friendship, even when that goal is an infinite end. The friendly person knows how to be a good friend in any given situation, and acts from that knowledge. The merely striving one knows, at best, how to actualize friendship in herself, where she hitherto lacks it. (This does not mean that the striving person's aim is merely "becoming" friendly, an expiring aim. She aims to *be* friendly, not merely to become so. Becoming friendly may, of course, be another aim of hers.) This is why even good instrumental reasoning in pursuit of an infinite end, mediated by true beliefs about the means, does not amount to having and acting from friendship.[93]

The same points apply, *mutatis mutandis*, to other character dispositions and the know-how they embody, such as open-handedness, honesty or courage on the one hand, and malice, greed, dishonesty and cowardice on the other. There would of course be an additional strangeness in attempting to conceive of, say, cowardice as an infinite end that one tries to pursue successfully in being a coward. And it may seem equally strange to conceive of cowardice as a case of "knowing how" to go on by the lights of cowardice. But I think this latter strangeness is only due to a mistaken temptation to conceive of knowledge-how in

---

[92] I take the idea of self-constitution from Korsgaard 2009, though I put it here to a use she might object to.

[93] Might it be that what distinguishes the friendly person from the striving one is aiming at the infinite end of *maintaining* friendship? But while it would certainly be strange to think of the striving person as aiming at maintaining the friendship she acknowledges she lacks, it is nonetheless possible for one to aim at maintaining one's friendly disposition while lacking such a disposition: one might be under the misconception that one has it. (Such a misconception might be linked to a misconception about what friendship amounts to.) And again, it will not help to suggest that one's aiming at maintaining friendship must not be based on such a misconception, for this just amounts to saying that the friendly person must have friendship as well as the aim. What it is to have friendship is still left in the dark by the response.

terms of knowledge of means to ends; and pursuing the end of cowardice *does* seem a strange undertaking. (Though it is not, perhaps, impossible.) A different way of thinking about the know-how involved in practical reasoning allows us to dissolve the strangeness, as I explain below.

Remember that dispositions of practical reasoning are just dispositions to take certain considerations, *xyz*, as one's reasons for certain actions, φ, in certain situations. Whenever one transitions in thought from potential grounds *xyz* to an action φ, one takes an inferential step. Any such inferential step may be conceived of as proceeding in accord with a "rule," in the sense that there is a *way* in which one takes *xyz* as supporting action φ.[94] This "way" corresponds to the "form" of one's reasons for φ-ing, as one acts on them. But as our earlier discussion of know-how showed, one's knowledge how to go on is not knowledge *that* one is going on in that way. It is instead a disposition to act for reasons in a way that gets things broadly correct by the lights of the "rule" in accord with which one is reasoning. So no thoughts about cowardice need enter the cowardly person's deliberations, even in the "background." While talk of "correctness" in this context may still seem strange, it is merely conversationally strange. It is simply a way of recording the point that one could not be thought of as following *this* rule if one did not, in one's way of reasoning, approximately conform to the pattern it describes. (If the pattern is uncodifiable in an explicitly formulable rule, we may do better to say that our character concepts *name* rules or ways of reasoning, than that the relevant rules or ways can be in some other way "described."[95])

If the foregoing is right, then there are possible cases of agency, cases of acting "from" character traits such as friendship or cowardice, that Sophisticated Humeanism cannot account for. Sophisticated Humeanism is therefore worse than optional: it should be rejected as a general account of practical dispositions.

---

[94] And one may always inquire as to the goodness of such a rule; the constitutivist instrumentalist attempts to argue that the rule in accord with which one reasons must always be an instrumental one, and on this basis to argue that instrumental reasoning is good reasoning.

[95] On uncodifiability of character, where this involves uncodifiability of one's conception how to live, see esp. McDowell 1979: 57-71; though McDowell's concern with uncodifiability comes in as a concern about how to think of the "major premise" of a practical syllogism. Cf. e.g. McDowell 1996 on the "blueprint" picture. My thinking about practical rule-following, character, and know-how has obviously been influenced by McDowell's work in myriad ways.

Sophisticated Humeanism initially entered our discussion as a way to support Practical Paralysis. Is there any other way to support Practical Paralysis? I consider this question below, arguing for a negative answer. For now, notice the consequences if the answer is "no." Insofar as the way one deliberates corresponds to the form of one's reasons, the friendly agent acts for reasons of friendship, best characterized as such, not for instrumental reasons. Likewise for the cowardly, the obsequious, and the honest. If acting "from" character is possible, as it seems to be, then instrumental reasoning, and acting for instrumental reasons, is not a condition of agency. Not only this: the result is a tentative content skepticism about constitutivism. For suppose, as seems to be the case, that e.g. the friendly and the malicious person are each intelligible as agents. And suppose, further, that one's having friendship does not depend on one's being malicious, nor vice versa. More generally, it seems, having virtue does not depend on having vice. These seem like reasonable assumptions. And now, suppose further that, as seems to be the case, there could be thoroughly virtuous agents, as well as thoroughly vicious ones. Since these agents' respective character dispositions are so different from each other, and since those dispositions cannot be analyzed in instrumentalist terms, it seems likely that the two agents need not share *any* practical dispositions with each other.[96] If they need not, then content skepticism about constitutivism follows. There is no particular disposition of reasoning one must have, no specific "rule" one must follow in deliberating and acting for reasons, to be intelligible as an agent. So there are no "constitutive norms" of agency. If authoritative practical norms exist, we cannot derive them from the conditions of agency.

So is there any other way, besides Sophisticated Humeanism, to support Practical Paralysis, or something like it? One might charge that the entire way we have been proceeding in assessing constitutivist instrumentalism is mistaken: we have been too focused on the moral psychology of what is in the "background" of one's reasoning. What makes our reasons for action instrumental in form is not

---

[96] Someone might suggest that there is some other disposition, besides acceptance of M/E, that the virtuous and the vicious must share. Indeed, we might think that vice is somehow merely a distortion of a disposition towards virtue that all agents have, or a failure of that disposition to properly manifest itself. This is Korsgaard's 2009 view: the relevant disposition is the disposition to follow the categorical imperative. In not directly arguing against such putative non-instrumental constitutive norms, this chapter's case for content skepticism about constitutivism remains tentative.

some moral-psychological fact about the allegedly necessary involvement of a background end-desire associated with a disposition to follow M/E; but rather, just the fact that action itself, as an unfolding event in the world, is articulated along means-ends lines.[97] The fact about the nature of agency that supports instrumental norms is just that whenever one does anything, ψ, one is always doing it *by* doing something else, φ.[98] One is always taking means to ends, and so one is always acting for instrumental reasons.

In response, however, it is hard to see how facts about what counts as doing what could on their own support claims about what, in particular, one should do. *So what* if I am not doing anything that counts as ψ-ing, but instead am doing something else? Surely most actions I could perform are such that I am not now doing anything that counts as doing them. It is hard to see how this entails that I am violating a norm. The constitutivist instrumentalist must say why the fact that ψ-ing is done by doing something else, φ, matters to *what, in particular,* I should be doing. And to accomplish this, it seems that she must retreat to an appeal to moral psychology: to some version of the claim that I should φ not only because ψ-ing consists in φ-ing, but also because *I am in fact doing, or trying to do, ψ.*[99]

One might concede this point, but nonetheless insist that the fact that action itself proceeds in stages or through some means implies that, no matter what other rules one might follow, one must always at least follow M/E. For no matter how one settles on some action, ψ, and no matter what one's reasons for doing so are—whether these be reasons of friendship or instrumental reasons—one must still heed M/E in order to actually articulate one's actions in the world. To articulate one's actions, one must choose some means, φ: and since choice of means is necessarily made in light of a belief that the means help one, somehow, to do whatever, ψ, one is aiming to do, instrumental justifications must at least be *among* those that agents must be responsive to. Thus all agents must accept M/E, whatever other rules they accept.

---

[97] This objection is inspired by Vogler 2002, though I am not sure she would make exactly this objection.
[98] Perhaps "atomic" or "basic" actions are an exception (Vogler 2002: 134). But I assume, with Vogler, that most actions of interest to us are not "basic."
[99] Though of course, that I am ψ-ing is not a *merely* psychological fact about me.

This objection supposes that the fact that action involves taking means to ends implies that there is a corresponding division of labor among the practical rules one accepts and acts through. Friendship, for instance, may get one to aim to relieve a friend's distress, but it cannot move one to actually do anything to relieve it. For that, a separate M/E principle is required. If this were right, M/E would indeed be a condition of agency, even if no particular rules governing ends are. But we should not accept the bifurcationist assumption that acceptance of a separate instrumental rule is required to get one to take means to ends, to articulate one's actions in the world. I show why in two steps: the first concerns identifying means, and the second concerns taking them.

First, it is clear that a separate M/E principle provides no guidance concerning how to *identify* the best means. The best person to appraise the appropriateness of means to the ends that one's friendly disposition makes one adopt is the friendly person herself. Sometimes the appraisal of appropriateness is tightly entwined with appraisals of efficacy; as e.g. when considering the best means to cheer up one's friend. At other times, the connection is less tight, as perhaps when considering the specifics of how to pull off a surprise party. But in considering how to pull something off, one need never be responsive to considerations of *sheer* efficacy, in the sense in which an option's efficacy might be appreciable from outside the perspective of the friendly person. As Anscombe (1995: 32) notes, one way to roast the pig is to burn the house down: that would get it done! But even if I know that burning the house down would roast the pig, I need not view this as even defeasibly recommending burning the house down. I may view it as unthinkable, since it would leave my friend without a home—a reason that gets a grip on me because of my friendly disposition.

So the friendly agent chooses φ-ing not just because it is *a* way of doing something, ψ, that she is doing, but rather because she implicitly conceives of φ-ing as an *appropriate* way to ψ by the lights of friendship. (More probably, the option of burning the house down never even appears practically salient to her.) One's choice of means is guided, in inextricable combination with considerations of efficacy, by implicit attunement to putative values other than instrumental value. Moreover, if, in attending to what ψ-ing would involve, one comes to view all the possible means to ψ-ing as odious for some reason or other,

one may intelligibly give up ψ-ing as a thing to do, even as one previously saw positive reasons to ψ and no reasons not to ψ. The friendly person need not take herself to have any reason to take the odious means merely because they are means to something she was going to pursue, or even saw independent reason to pursue.[100]

In this way, the friendly person's knowledge how to go on by the lights of friendship, in any given situation, issues in both correct ends and correct beliefs about means—correct, that is, by the lights of friendship.[101] But notice, crucially, that even though one's practical thought is in part *about* ends and means, at the stage of considering how to do something one has settled to do, this does not imply that one's practical thought is instrumental in *form*—any more than one's "foreground" considerations' *not* being about ends and means, but instead about (say) duty or friendship, implied that one's reasoning is *not* instrumental in form. Thought about means and ends is not thereby thought that is instrumental in form.

However, second, once one has identified the preferred means, is a separate M/E principle at least needed to get one to *take* these means? One might think that, although the friendly person's knowledge how to go on by the lights of friendship issues in correct beliefs about means, she cannot *exercise* this know-how, and so actually take those means, except by following M/E. This would mean that one's friendly disposition cannot issue in action without the mediation of a disposition to follow M/E. If this were right, then one could only ever follow other rules by following M/E. But it is hard to see why we should adopt this view. Citing a separate M/E principle cannot explain what it is to exercise one's know-how in following practical rules. In order to follow M/E itself, one must know how to go on by its lights, and exercise this knowledge. The idea that M/E is an intermediate rule through which one follows all of one's other rules would lead to yet another rule-following regress (though a different one than Dreier's):

---

[100] I thank John McDowell for pressing me to explain more clearly why one need not take sheer efficacy in the service of an end as at least a *pro tanto* reason in an option's favor.
[101] They might of course also be correct, or incorrect, by some other standard.

to exercise know-how, one must follow M/E, but to follow M/E, one must exercise one's knowledge how to follow it. Practical rule-following would be impossible.

So the idea that we must accept a separate M/E principle does not help account for the idea of efficacious practical rule-following. It merely compartmentalizes that idea. We might suggest that, in a way, all that M/E tells us is to be efficacious, not idle. Perhaps it just says: "Exercise your know-how!" But the fact that a rule tells us to be efficacious, not idle, does not account for what it is to efficaciously follow even that rule. Indeed, if all that M/E instructed us to do is to exercise our know-how, it would no longer be a rule one could independently follow. It would instead be a summary or codification of what goes on in any case of rule-following: one exercises one's know-how. Nor could M/E, thus understood, be normative. For whenever one manages to follow practical rules at all—whenever one manages to deliberate, or act for reasons—one exercises one's know-how in *some* way. And if there is no possibility of discord with M/E while deliberating and acting for reasons, then M/E cannot be a practical norm.[102]

These points about efficacy in practical rule-following are just ways of articulating the familiar point that there must be a way to follow rules *immediately*, without the mediation of a further rule. If these points are correct, then it looks to be possible for one's friendly disposition itself to be what moves one in both choosing and articulating one's actions in the world. In exercising her know-how, the friendly person chooses courses of action according to how they seem appropriate by the lights of friendship. And if action itself proceeds in stages and through some means, then in being moved to act, she is moved to actually go through the stages and take the means that she has already identified as appropriate. This is what it is for her to articulate her friendly actions in the world. The point is that in doing all of this, one *just is* following the "rule" of friendship, doing what one implicitly conceives of as appropriate in its

---

[102] What if M/E tells us to exercise our know-how *correctly* by the lights of whatever other rules we are following? Then one could go wrong by M/E's lights. But there would still be no possibility of *independent* discord or accord with M/E. It would be a mere side-effect of following our other rules correctly or incorrectly that we also "follow" M/E correctly or incorrectly. M/E would thus add nothing to a system of practical norms. In thinking about the relationship between M/E and exercising know-how, I was much helped by objections from Kieran Setiya and two anonymous referees.

light. (Of course, this is not an independent account of what it is to exercise know-how. Nonetheless, the general point remains that citing a separate M/E rule cannot account for it.)

I argued earlier that acting "from" friendship or other character traits cannot be analyzed in Sophisticated Humean terms, because the friendly person's know-how cannot be captured in terms of knowledge of means to ends. On the other hand, I have just acknowledged that the friendly person's knowledge how to go on *issues* in knowledge of what means are appropriate to what ends in any given situation, by the lights of friendship. Knowledge of means to ends is thus part of the friendly person's thinking; but only as a *topic* when considering how to achieve what one is trying to achieve, not as giving one's practical reasoning instrumental form. I conjecture that it is precisely the fact that many actions require means-ends articulation, so that deliberation that issues in those actions must be in part *about* means and ends, that is most responsible for the impression that accepting M/E, or something like it, is a condition of agency. But while it is true that the fact that φ-ing would count as ψ-ing, or would help one to ψ, must be at least an aspect of what recommends φ-ing to one and makes one choose it, this has the status of a "foreground" consideration in the person acting "from" friendship. It does not imply that in being moved to φ as a way of ψ-ing, one's reasons for doing so are instrumental in form.

If these points are right, then we can diagnose the source of the impression that means-ends thought is in some way essential to agency, without thereby admitting Practical Paralysis. We have seen that neither Sophisticated Humeanism, nor the means-ends structure of actions as unfolding in the world, nor the notion of exercising one's know-how in practical rule-following, vindicates Practical Paralysis. It is hard to see what else could.

There is a final protest to all of this. Even if we cannot give acting "from" character an instrumentalist analysis, not all contexts of action are occasions for acting through efficacious character traits. We might even suspect that there could not be agents who *only* act through their various character traits, and not at all in pursuit of personal projects whose apt realization has nothing to do with their character. In response, I am not sure what to say about the importance of personal projects to an analysis of the nature of agency. But it seems that even when it comes to the pursuit of seemingly minor projects,

one's various character traits can, in part, guide one's choice of means. This was the case in our earlier example of the friendly agent preparing to roast the pig.

I started §2.4 by asking how best to think of friendship and other character dispositions as moving us to act. A summary way to view my response to this question is that I have been trying to articulate a rationale for the basic conviction that it is wrong to think of friendship as in itself idle, able only to select ends—or perhaps, in addition, to impotently specify means. One who has supposedly already identified the best way to help but nonetheless fails to come to her friend's aid is not just instrumentally irrational: she is only dubiously someone who has friendship. Friendship is precisely a practical disposition. In deliberating as the friendly person does, and seeing the reasons the friendly person sees, one is moved to actually act as the friendly person acts. Putting thought about means and ends into its proper place in an account of practical rule-following allows us to do justice to this conviction.

2.5 CONCLUSION

There is an often-implicit tendency in much contemporary anti-instrumentalist work to structure one's opposition to instrumentalism in terms that are favorable to the instrumentalist's bifurcationist claim that an independent instrumental norm governs the pursuit of means to ends, so that the only possible role left for other putative practical norms is that of governing ends. Such opposition to instrumentalism has to either argue that specific end-governing principles are also essential to agency, or else take issue with the entire constitutivist strategy in some way different from what I have done. But the bifurcationist way of framing the debate obscures the possibilities of agency that I have been trying to articulate. If I am right, the correct anti-instrumentalist response to constitutivist instrumentalism is not to add to the requirements of agency, as some have tried to do, but rather to adopt a more liberal view of those requirements. Agents as such need not even follow an instrumental rule of reasoning, nor act for instrumental reasons. There is therefore no good constitutivist argument that they should. Instead, agents may be disposed to reason in

various different ways, picked out, for instance, by our character concepts. This leaves the field wide open as regards putative practical rules one might intelligibly accept. So open, in fact, that we should be skeptical about the claim that the conditions of agency can yield insight into the content of practical reason. If the suggested liberalism withstands scrutiny, constitutivism fails.

3.0 AUTONOMY AND CONTINGENCY

3.1 THE TOPIC AND WHY IT MATTERS

It is fairly commonly held that agency is essentially autonomous in some sense. Autonomous agency is sometimes glossed as *self-governing*, or *self-determining*, or (perhaps more controversially) *self-legislating*.[103] It is also often said that an autonomous agent's actions must be *attributable* to their agent; or that agents must be the *authors* of their actions;[104] or that, when an agent acts, her bodily movements must be caused by the *agent herself* in some way, rather than by some "alien" causes operating on her, or even "in" her (as perhaps with bodily twitches caused by random neuronal firing).[105] The agent's actions must be *her own* to really be her *actions*, rather than just movements of her body to which she is some sort of a passive spectator.[106]

I will take it as a given here that there is some truth in these sorts of pronouncements: that there is indeed an essential aspect of agency which we may call its 'autonomy', an aspect that helps to separate exercises of agency from mere happenings involving agents. What I would like to try to understand is exactly what this means in the context of an account of agency—in particular, what it means for an account of the agency of thinking beings, beings who are rational in the sense contrasting with *non-*

---

[103] For autonomy as self-governance or self-determination, see e.g. Korsgaard 2009: 69, 106-7 *et passim*, Watson 1987: 191-6, Ekstrom 2005: 155. For autonomy as self-legislation, see Korsgaard 2009: 108, 127.
[104] For autonomy as "authorship," see e.g. Velleman 2009: 130 *et passim*. For autonomous actions as attributable to their agents, see e.g. Korsgaard 2009: 100-104.
[105] For autonomy as defined in contrast to determination by "alien" causes, see e.g. Frankfurt 1975: 48, Watson 1975: 26, and esp. Korsgaard 2009: 89-90, 106.
[106] For autonomy as action that is "one's own" in some special sense, or as action on motives that are "one's own" in some special sense, see e.g. Frankfurt 1977: 59, Bratman 2003, and differently, Watson 1975: 26-30.

rational. Rational agents in this minimal sense act through exercising their rational capacities. By the rational capacities relevant to agency, I understand the capacity to act for reasons, and to deliberate towards action, whatever these amount to. (To be rational in this minimal sense, to have and exercise rational capacities, is not yet to be rational in the further, commending sense contrasting with *ir*rationality; though the two senses of rationality are of course connected in that only minimally rational agents are so much as capable of irrationality.) My question is: What is autonomy, as an essential characteristic of acting for reasons and of reasoning towards action? What is the autonomy of rational agents, just as such, in the minimal sense of 'rational'?

In asking this question, I acknowledge that there may be senses of 'autonomy' in which autonomy is *not* an essential feature of rational agency as such, but rather just an aspect of a form of agency that we especially value. Whatever such especially valued forms of agency might amount to, I wish to put those aside here. My focus is just on the sense in which rational agency as such is essentially autonomous. But if this sense of autonomy is potentially narrower than other, especially valued senses of autonomy, then why is it worth inquiring into?

The topic is worth inquiring into not only for the sake of understanding agency, but also because of its connection to another pressing philosophical topic: the nature and content of practical norms. By "practical norms," I just mean norms governing deliberation and acting for reasons. A practical norm tells us what it is to deliberate *well*, or as one *should*, and what it is to act for *good* reasons. Some philosophers inspired by Kant have recently argued that the fact that agents must be autonomous helps to yield not only an account of agency, but also (i) an account of what the practical norms binding agents are, as well as (ii) an explanation of those norms' authority. The idea is that we can derive "ought"-claims, about how we should reason and act, from "is"-claims, about what rational agency just as such is.

For instance, Christine Korsgaard argues that implicit in the fact that rational agency is autonomous is a commitment to Kant's categorical imperative (CI).[107] CI is a rule of practical reasoning that tells us to act only on maxims whose form is universal:

> **CI** Act only in accordance with that maxim through which you can at the same time will that it become a universal law.[108]

Exactly what having "universal" form means will be explored below. But Korsgaard's main thought is that one cannot act autonomously except through reasoning in accord with CI;[109] and that CI is authoritative because of this (2009). Differently, David Velleman argues that autonomous agency requires one to be disposed to act on considerations that display the action to be chosen as what is "most intelligible for [one] to do" in the light of one's particular "circumstances, attributes and attitudes" (2009: 132-33). And Velleman holds that the considerations that display an action φ as "most intelligible" for its agent in this way thereby qualify as authoritative practical reasons in favor of φ-ing (2009: 128-33).

Each of these arguments attempts to derive practical norms from the metaphysics of agency. Call arguments of this general sort *constitutivist* arguments. There are two main stages to such arguments. First, there is the claim that deliberating and acting through some specific disposition, a disposition to reason in accord with a rule R, *just is* what exercising autonomous agency is. For Korsgaard, the relevant disposition is that of following CI. For Velleman, it is the disposition to follow something like the rule "Do whatever it is most intelligible for you to do, given your circumstances, attributes, and attitudes."[110]

---

[107] Korsgaard also thinks that implicit in the fact that rational agency is *efficacious* is a commitment to Kant's hypothetical imperative (HI) (2009: 59-72, 81-90). I comment on this briefly below.

[108] Kant, *Groundwork* 4:421.

[109] To be precise, Korsgaard views "acting on a rational principle" as not necessarily involving any explicit "step-by-step process of reasoning," for a principle can be "deeply internalized" so that we simply "*recognize* the case as one falling under the principle, where that is a single experience" (2009: 107). Nonetheless, her explication of the deep structure of practical thought proceeds on the basis of the idea that it involves reasoning in accord with a principle. I follow her here, understanding acting for a reason as exhibiting the same basic structure as reasoning in accord with a principle: when we reason, we make a consideration, a "premise," our reason for something else, a "conclusion." Whether that process is explicit or implicit, or slow or fast, is not as important as its basic metaphysical structure, which is shared in each case. (More on this later.)

[110] §5 will consider what "intelligibility" amounts to for Velleman, and whether we should consider it a constitutive norm of practical reason.

The constitutivist's claim is that you cannot deliberate and act autonomously at all except through the disposition in question; but when your actions *do* issue from exercising the constitutive disposition, they thereby count as exercises of autonomous rational agency. There is room for failures here, since the constitutive dispositions might be imperfectly manifested. In following a rule of reasoning, I may do so imperfectly, with performance errors. But just as I can still count as speaking English even when I make some mistakes, likewise my bodily movements can count as exercises of the dispositions constitutive of autonomous agency even when I make some mistakes in reasoning. The key is just that my movements issue from dispositions constitutive of autonomous agency, however imperfectly those dispositions may be manifested on occasion.[111]

The second stage in constitutivist arguments is explaining the transition from this purported metaphysical "must," about the dispositions constitutive of autonomous agency, to a normative "must," about which putative practical norms are genuinely authoritative. How can we derive an "ought" from an "is"?[112]

In a way, the details of the derivation will not matter for my purposes here. What I will be arguing for in the bulk of this chapter is that there *is* no practical disposition that is constitutive of autonomous agency, from which to derive practical norms. Autonomous rational agency can be exercised to the full through dispositions that are *contingent* for agency as such. That is my thesis about autonomy. But it will be worth seeing in outline how the constitutivist derivation might go, so as to see why its premise is even worth engaging. If the derivation is bound to fail, then nothing we say about autonomy can affect the success of constitutivism. I think that, to the contrary, there is a way to spell out a

---

[111] There are vexing questions concerning what exactly the difference is between, on the one hand, imperfectly following one rule, and on the other, perfectly following some subtly different rule. Further, what is the difference between being disposed to follow R, as R really is, and being disposed to follow R only to an approximation, so that your disposition itself is imperfect? These are problems for a view of rule-following on which rule-following is a matter of manifesting dispositions. The problems concern finding criteria of identity for the dispositions we have and manifest, where the dispositions cannot simply be read off one's behavior. (Cf. Boghossian 2008a: 130.) As far as I know, there are no fully satisfactory solutions to these problems yet. Nonetheless, the disposition view, with all of its problems, remains the most promising extant view of rule-following. (For compelling criticism of a competing "Intentional View," see Boghossian 2008a: 127-8; I discuss this in ch.2 of this dissertation, at the end of §2.2.)

[112] Cf. Hume's famous remark at *Treatise* 3.1.2:27.

constitutivist derivation so that it works; and that consequently, what we say about autonomous rational agency is of crucial importance to anyone who cares about what the norms are by which we should guide our practical lives. So how does the derivation of "ought" from "is" go?

One sort of "ought" is fairly easy to find in this area. If R is a rule of reasoning that one can follow either correctly or incorrectly, it has what we might call 'rule-internal normativity'. Whenever one follows R, one can either meet or fail to meet norms that apply to one, simply in the sense that R says something about how one should proceed in one's situation, and one might go either right or wrong by R's lights. In this sense, if any exercise of agency involves reasoning in accord with some rule or other, then any exercise of agency puts you under norms that apply to you—the norms of rule-internal normativity.[113]

But of course, rule-internal normativity is not yet *authority*. Any old rule can "apply" to you in the sense of having something to say about how you should deliberate and act in your situation, such that your actions can then be either right or wrong by the lights of the rule. Consider the rule Mother Says: a rule that tells you to do whatever you believe your mother would want you to do in your situation. This rule has rule-internal normativity. It makes a claim about how you should go about deciding what to do, and you could go wrong by its lights. But any old rule that we might make up can make claims about what we should do and how we should reason. Indeed, there are an infinity of such rules. What we want to know, in wanting to know what the norms are by which we should guide our practical lives, is not just what this or that rule claims about how we should reason and act. What we want to know is just: how we should reason and act. Which rules' claims should we really listen to and accept? Which rules' claims have authority?[114]

An exactly parallel distinction between rule-internal normativity and authority applies to the case of theoretical reason. For instance, both induction and counter-induction make claims about how we

---

[113] In fact, a rule can "apply" to you, in the sense of claiming something about how you should go on in your situation, without your even following that rule; and your actions may be wrong by the lights of the rule.
[114] Compare Foot 1972, esp. 160-1, 164-7, on rules' "application" vs. their authority.

should go about forming our beliefs. A gambler who follows a rule of counter-induction, a rule telling her that the future will diverge from the past in exciting ways, might follow counter-induction either correctly or incorrectly. Counter-induction has rule-internal normativity, just as induction does. But this does not settle the question which rule's claims the gambler should really listen to and accept.

Clearly the question concerning the "ought" of practical norms is not just a question about rule-internal normativity, but rather, a question of authority. Rule-internal normativity is plausibly a *prerequisite* of a rule's authority. Since an authoritative practical norm tells you, authoritatively, how you should reason and act, any putative practical norm must at least have something to say about that topic. But rule-internal normativity does not yet amount to authority. How can the constitutivist derive the authority of a putative practical norm from the metaphysics of agency?

Very roughly, we can think of the derivation as follows. The constitutivist converts the rule-internal normativity of a rule to authority, *via* the metaphysical claim that following that rule is constitutive of agency as such. How does this work?

Suppose rational agency *just is* a matter of following some specific rule(s), R, however imperfectly, as the constitutivist's "is"-claim states. If that claim is true, then *deliberating and acting for reasons well, tout court,* must be a matter of *following R well*: that is, a matter of following R while meeting its rule-internal norms. That follows by a plausible principle of substitutivity: necessarily co-referring terms are substitutable *salva veritate* (in non-oblique contexts). (By this I mean that, if t1 refers to A and t2 refers to B, and □(A=B), then t1 and t2 can be substituted for each other in statements without changing the truth value of the statement, so long as the statement is not the content of a propositional attitude.)[115] The idea is simply that if it is in the metaphysics of agency that following R *just is* deliberating and acting for reasons, as the constitutivist's "is"-claim states, then the terms 'following R' and 'deliberating and acting for reasons' are necessarily co-referring. Hence following R well, or as one

[115] I say "necessarily co-referring terms," rather than "co-referring terms," to forestall the objection that contingently identical things (such as a statue and the lump of clay that composes it) do not make true the same modal judgments. (I thank Kieran Setiya for this objection.) The constitutivist's metaphysical claim about agency is one of necessary identity: necessarily, exercising rational agency just is following R.

should, *just is* deliberating or acting for reasons well, or as one should. In following R well, one meets the demands of R's rule-internal normativity; but one *thereby* also deliberates and acts for reasons well, or as one should, *tout court*.

And now, what more could practical norms require of one than deliberating and acting for reasons well, or as one should, *tout court*? The answer seems to be: nothing. Practical norms just are norms telling us how we should deliberate, and what reasons we should act on and how. If one deliberates well, *tout court*, then one deliberates as well as one can, *qua* rational agent. We could hardly ask more of rational agents. In this way, the gap between R's rule-internal normativity and its authority is closed by the constitutivist's "is"-claim, that agency just is a matter of following R.

That is a very quick sketch of the derivation, and there is certainly much more to say about it. I have addressed it in more detail elsewhere.[116] But I do think that an argument of the basic form just sketched works, given the appropriate "is"-claim. If that is so, then engaging with the constitutivist's "is"-claims is worthwhile. In fact, anyone who cares about the norms by which we should guide our practical lives should also care about the proper understanding of autonomous rational agency. For *either* some constitutivist "is"-claim is right, and so yields authoritative practical norms; *or* no constitutivist "is"-claim is right, and the content and justification of practical norms cannot be settled by doing philosophy of action. To establish either result, though, we are forced to engage with the question what rational agency just as such is—and so, with the sense, if any, in which rational agency as such is autonomous.

There is a caveat to what I have said thus far. The constitutivist argument claims to derive practical norms from truths about what rational agency as such is. And the fact that rational agency is autonomous is just one fact about rational agency. There is another seemingly essential aspect to rational agency—namely, its *efficacy*. We cannot act through exercising our rational capacities unless these capacities are efficacious in actually moving us. And we might think that efficacy imports its own constitutive norms. Korsgaard thinks it does: she thinks that just as autonomy consists in following CI,

---

[116] Chapter 1 of this dissertation.

efficacy consists in following a hypothetical imperative, HI.[117] Accordingly, she thinks that deliberation proceeds in two stages or has two aspects—formulating a maxim through following HI, and testing that maxim for universal form, through CI. Each stage is necessary, individually insufficient, but jointly sufficient for action.[118] However, these complications will not matter for my purposes here. The basic idea of the constitutivist argument remains unchanged. One's actions issue from an exercise of one's rational capacities if and only if they are manifestations of the disposition or dispositions of reasoning constitutive of rational agency as such; and the metaphysically constitutive status of these dispositions makes the claims of the rules they are dispositions to follow authoritative for rational agents as such. I have addressed efficacy as an aspect of agency elsewhere.[119] Here my focus will just be on autonomy.

To investigate autonomy, I will proceed by investigating the central concepts concerning autonomous rational agency. In particular, I will focus on the ways in which Korsgaard and Velleman employ those concepts to motivate their respective constitutivist "is"-claims. The reason for focusing on these authors here is just that they are the authors who have most extensively pressed the idea that autonomy imports constitutive norms. But in arguing against their conclusions about what autonomy entails, I mean to show, quite generally, how a proper understanding of autonomy is compatible with thinking that autonomous rational agency can be exercised through contingent practical dispositions. There is no metaphysical "must" of the sort the constitutivist needs, from which to derive practical norms. There are of course necessary aspects to agency—for instance, the fact that agency is necessarily autonomous. But the point is that there is no putative practical norm such that autonomous agents necessarily act through following *it*, however imperfectly, so that the constitutivist derivation could have a chance of showing that following that norm perfectly must amount to deliberating and acting for reasons well, *tout court*.

---

[117] Other authors, e.g. Smith 2009, 2010 and Dreier 1997, have similarly argued that efficacy requires following some sort of an instrumental norm.
[118] Korsgaard 2009: 59-72, 81-90. I discuss this in more detail in chapter 1, §1.3.2.2.
[119] In chapter 2 of this dissertation.

§§3.2-3.4 focus on Korsgaard, §3.5 on Velleman. §3.2 considers the concept of *self-determination*. §3.3, the longest section, inspects the idea that autonomy requires a degree of *unity* in the agent's psyche in a way that imports constitutive norms. The same section also considers the notions of *identifying with* one's maxims or motives, and of the *universality* of one's maxims, both of which are notions connected to that of unity in Korsgaard's constitutivist argument. §3.4 considers *self-legislation*; §3.5 examines the concepts of *self-knowledge* and *self-understanding*, and Velleman's use of these notions in his constitutivist argument. To be sure, the various notions I investigate are not entirely separate. But taking each in turn, thereby building up our understanding of autonomy, will help to isolate exactly where constitutivist views of autonomy go wrong.

3.2 SELF-DETERMINATION

To see why autonomous agency is intelligible without constitutive norms, start from the intuition that autonomous action is self-determined action. What is it for me to determine myself to act, rather than for something else, something alien or external to me, to determine my behavior?

To work towards an answer, let us begin with a basic insight of Korsgaard's. She notices that there is a perfectly generic notion of self-determination, under which the self-determination of rational agents falls as a species. According to this generic notion, an agent (of whatever sort) is self-determined when she is determined to act by the "principles of her own causality," whatever they are. What does this mean? We can best explain through examples. For instance, non-rational animals have their own principles of causality in the sense that they have *instincts* that determine what cues in the environment the animal responds to and how. An animal's instincts determine, first, which features of objects the animal views as *incentives*—as somehow attractive or aversive—and second, exactly how the animal

responds to these incentives, what it does in their light.[120] The cat is attracted to the small scurrying thing in the garden, and her response is to prowl, adjust her footing, and then pounce. The central idea is that the animal just *is* her instincts, in some sense; and when the animal acts from her instincts, she acts from her essence or "form."[121] When the animal's behaviors issue from this form, they are self-determined.

The same generic notion of self-determination applies also to rational agency. Since the activities through which rational agents exercise their agency are the activities of reasoning towards action and of acting for reasons, it must be that self-determined action for rational agents is action that issues from these activities. Autonomous rational agency is a matter of being moved, not by any old impulse, or in any old way, but rather, by your making some considerations *p* your reasons for doing φ, and acting on those reasons. That, in very general terms, is the essence or "form" of rational agency as such. It does not matter whether your making *p* your reason for φ-ing involves a long chain of deliberation, or instead just a one-step inference, a quick transition in thought from grasping *p* to doing φ. In either case, when rational agents act from this "form," they are self-determined. When they do not, their bodily movements are "alien" to rational agency.[122]

In fact, this generic notion of self-determination seems to apply across the board, even to plants. In this generic sense, an oak tree determines itself in growing its acorns, but not in being hit by lighting. We might say that in being hit by lightning, the oak tree is *passive* with regard to what happens to it, whereas in growing its acorns, it is *active*, *qua* the kind of being that it is.[123]

Korsgaard's generic notion of autonomy as self-determination, as explicated so far, seems right to me, including in its application to rational agency. How could autonomy in acting for a reason fail to be a

---

[120] Korsgaard 2009: 104.
[121] For self-determination as determination by one's "form" in the sense of essence, see Korsgaard 2009: 107, *et passim*.
[122] Are there autonomous actions, exercises of rational agency, that do not involve acting for a reason? It seems to me that G.E.M. Anscombe is right in saying that acting for no reason cannot be the central case of action (1957: §20). But I will not argue for this here. In any case, even if it were the central case, it is hard to see how a constitutivist could derive claims about good reasons for action from it.
[123] I am thankful to John Carriero for the oak tree/lightning example. He tells me that the conception of self-determination, and of activity and passivity, explicated here is Spinozistic.

matter of being moved by one's making some consideration, *p*, one's reason to φ? But let us specify that notion further. What I say next is not something that Korsgaard says about self-determination. Nor is it anything with which, I think, Korsgaard would disagree. Nonetheless, it will be crucial to the way in which I will argue against Korsgaard shortly.

Whenever any actual animal is governed by instinct, there must always be some *determinate* instincts in play. An animal can't be governed by instinct, just as such. Typically we discover different determinate instincts in play in different species of animal, or in different types of animal as typed by their role in their biosphere. Predators and prey have quite different sorts of instinct, for example. But the point is that there has to be *some* determinate instinct or other, for an animal to be determined to act through her instincts.

In the same way, it seems that rational agents cannot simply be governed by the activity of deliberation and acting for reasons, just as such. Whenever you deliberate, you must always deliberate in some *way* or other, in accord with some rule or other. There must always be some determinate shape to it. And correspondingly, whenever you make a consideration *p* your reason to φ, there has to be some *way* in which you relate your reason to the action, some way in which you take *p* to be relevant to the question whether to φ. You cannot simply make absolutely *anything* your reason, willy-nilly—you cannot just take a random consideration and a proposed action, and decide to make the consideration your reason for the action—without having some implicit conception of how that sort of a reason could be a reason for that sort of an action. For example, you cannot decide to make pesto just because the Monongahela is high—not unless we can tell some back story that illuminates the *way* in which that sort of thing, the water level, could be somebody's reason for making pesto.[124] There has to be some determinate publicly intelligible pattern, it seems—some "rule," loosely construed—that you are following, and that makes sense of *how* it

[124] Setiya 2010: 97 makes a similar point in a different context. Setiya argues that if your reason is not antecedently intelligibly related to the action it is supposed to be your reason for, then we cannot make the reason intelligible as your reason for the action by simply insisting that you think it is a *good* reason for the action. The implication is, I take it, that there has to be some other way of making the relation between your reasons and your actions intelligible. I return to this in §3.3.3.

is that you take *p* to be relevant to φ-ing in particular, so that your practical reasoning leads you to φ because *p*. It is only through following determinate rules of reasoning, in this loose sense of 'rules', that rational agents can exercise their specific kind of autonomy.

So the concept of self-determined rational agency is a determinable concept. And we do not have any actual exercise of rational agency in view until we have some particular determinate form of it in view—where a "determinate form" of rational agency is picked out by whatever rule of reasoning one follows in exercising one's agency. Whenever you reason, you reason in accord with some rule or other. And when you reason in accord with some rule, you reason. Rational agency cannot be exercised except through some determinate form of it, but whenever you have and exercise some determinate form of it, you are self-determined *qua* rational agent.

Although Korsgaard does not make the determinable/determinate distinction I just made, I doubt she would disagree with it. What she and I disagree about is which determinate forms of rational agency are possible. Korsgaard thinks that it is part of the essence of autonomous rational agency as such that the only possible determinate form of it is following CI (or CI+HI), however imperfectly. I disagree. Self-determination can be action through contingent determinates of determinable essences. That is what I will argue for.[125]

My argument for this will be in two stages. First, I will show why nothing in the notion of self-determination as such imports a need for conceiving of some determinate form of agency as the only possible one. That will complete this section (§3.2). But the argument is not complete until we consider the connection of the notion of self-determination to other notions in the area. I take up those other notions, and the ways in which they modify our understanding of what the self-determination of rational agents involves, in §§3.3-3.5.

---

[125] There are complications: If CI were essential to agency as such, then wouldn't it be part of the "determinable essence" of agency? On the other hand, may Korsgaard not regard "following CI" as itself capable of further, perhaps contingent, determinations? However we answer these questions, it will not change the main point of disagreement. Korsgaard must regard CI as being, itself, a determinate form of reasoning. And it is the claim that following CI is both sufficient and *necessary* for agency, so that agents cannot follow any *competing* (not subsidiary) rule of reasoning *instead*, that is supposed to show CI's authority.

For now, consider the notion of self-determination as such. Nothing in the bare concept of autonomy as self-determination, as explicated thus far, requires that the rules we follow be necessary. In the case of animals, it certainly seems that it must be possible for the instincts through which an animal acts to be contingent, without this jeopardizing the animal's autonomy in acting through them. Not all instincts are necessary for non-rational animals as such: after all, as I just said, predators and prey have some rather different instincts. But the contingency of a particular instinct, from the point of view of the "form" of non-rational animality as such, seems to be no reason at all to think that when a particular animal *acts through that instinct*, the animal's actions are less than self-determined. Just as a color can be a determinate color without being specifically *this* determinate color, likewise an animal can act through determinate instincts—as it must, if it is to act through its instincts at all—without acting through *this* determinate instinct. So long as it is on its instincts that the animal acts, her actions are self-determined. It does not matter whether the determinate instincts are contingent so far as the essential nature of the determinable *non-rational animal* is concerned.

A parallel argument applies to rational agency. Nothing in the bare concept of self-determination, as applied to rational agency, requires that the rules you follow in reasoning be themselves necessary. All that is necessary is that, in deliberating and acting for reasons, you follow some rule of reasoning or other, and that you (thereby) reason. If you reason in way R, you reason. If you reason in way R*, you reason. Either way, the actions that issue from your reasoning are self-determined in the sense that they issue from determinate forms of rational agency.

Of course, Korsgaard's claim is precisely that you *cannot* deliberate in any way except in accord with CI (or CI + HI). But the present point is just that nothing in the bare concept of autonomy as self-determination helps Korsgaard to argue for this claim. Nor indeed does the concept of self-determination help us towards any parallel constitutivist proposal. The point that self-determined action can be action through contingent determinates of determinable essences is perfectly general: if it stands, it is destructive to all constitutivist proposals.

One might object that each animal's instincts must be at least necessary, if not for non-rational animals as such, then at least for that particular type of animal. The cat would not be a predator any more if it completely lacked the instincts of predators. Perhaps it would not even be a cat any more. That may be so. But nothing analogous to this suggestion could help the constitutivist. The constitutivist is trying to derive practical norms from the metaphysics of rational agency as such. It may well be that particular determinate forms of rational agency each have their own essences: for example, of course you cannot be a CI-follower unless you follow CI. CI-following is the particular shape that self-determination necessarily takes in any rational agent *that is a CI-follower*. But that is trivial, and exactly analogous truths apply to any other practical rules we might care to propose. I cannot be a follower of Mother Says unless I follow Mother Says. That is a truth about the essence of being the particular determinate form of reasoning agent that I am in being a Mother-obeyer. What the constitutivist needs is more than this. What she needs is the claim that there is some determinate form of rational agency, some disposition of reasoning, that is constitutive of rational agency *as such*. Without this stronger claim, the constitutivist argument does not go through.

I have argued that nothing in the bare concept of autonomy as self-determination requires the strong claim that the constitutivist needs. Self-determination can be action through contingent determinates of determinable essences. However, in arguing for the claim that autonomy requires following CI, Korsgaard further connects the notion of self-determination to the notions of *psychic unity*, *identification*, and *universality*. The next section investigates these developments, arguing that autonomy remains consistent with the contingency of one's practical dispositions.

3.3 UNITY, IDENTIFICATION, UNIVERSALITY

3.3.1 Unity, synchronic and diachronic

According to Korsgaard, the fact that action is essentially self-determined means, in part, that action requires an agent, a *self,* from whose activity one's bodily movements issue. Self-determination is "determining yourself to be a cause" (2009: 72) of your movements. And, she writes,

> […] determining yourself to be a cause is not the same as being moved by something within you, say some desire or impulse […] operating as a cause. When you deliberate, when you determine your own causality, it is as if there is something over and above all of your incentives, something which is *you*, and which chooses which incentive to act on. So when you determine your own causality you must operate as a [unified] whole, as something over and above your parts […]. And in order to do this […] you must will your maxims as universal laws. [That is, you must deliberate in accord with CI.] (2009: 72)[126]

This passage does not yet claim to be an argument for the claim that operating as a unified whole requires following CI. It is merely an explanation of why autonomous agency requires the agent to be a unified whole, and a claim, to be argued for, that agential unity requires following CI. But even before getting to the argument for CI, we might already object that Korsgaard's idea of a unified self as something "over and above" one's various psychic elements sounds dubious. Korsgaard later clarifies, however, that the unified self in question is not supposed to be some mysterious *substratum*, a propertyless bearer of mental properties. Rather, it is just the idea of *having a constitution*: some constitutive procedure or procedures that characterize the way that one's various mental elements hang together and operate in concert. The constitutive procedures of a rational agent in particular are its rules of reasoning, rules that describe its way of acting in the face of the various impulses that it is passively subject to. The constitutive procedures thereby impose a sort of unity to the psychic zoo. Korsgaard agrees with Plato's city-soul

[126] Though the particular passage I quote does not include the word 'unified', I insert it since it is clearly Korsgaard's intent. See e.g. p. 59, which starts the chapter from which the passage quoted is lifted. There Korsgaard frames her chapter by saying that she will "explain how the principles of practical reason serve to unify the will, and how that makes them normative."

analogy: her thought is that neither Plato's city-state nor a soul can act as a whole unless its actions issue from its constitutional procedures.[127]

It is hard to assess the merits of the city-soul analogy. For one thing, the inhabitants of a city are themselves rational agents, whereas the various impulses in an agent's psyche are not. As a result, there are also bound to be differences in the criteria for something's being a constitutive procedure in a city-state *versus* in an individual soul. For instance, in the political sphere, it is rational agents, themselves autonomous, who institute constitutional procedures, and can topple them. There seems to be no analogy to this in the case of individual agency. Certainly Korsgaard herself would not want to think of the "constitutional procedures" through which rational agents can act autonomously as instituted and toppled by the very impulses that those procedures are supposed to control. This would still be a case of being ruled by impulses one is passively subject to, even if mediated by "procedures" that those impulses themselves create.[128]

Whatever the case with the city-soul analogy, however, there is surely some plausibility to the thought that an agent must have a measure of psychic unity—both synchronic and diachronic. We could not think of actions as attributable *to agents*, as opposed to just some inner goings-on, unless there was some sense in talking about agents as opposed to mere series of impulses that may or may not happen to initiate a bodily movement at a time.[129] Moreover, a single agent cannot perform any actions that it takes any time to perform—and that includes most, if not all actions we ever do—unless that agent also persists through time. So there is some need to see agents as more-or-less unified subjects, synchronically and diachronically. Given the implausibility of a *substratum* view on the one hand, and the plausibility of the thought that rational agency always takes place through following some determinate rules of reasoning,

---

[127] 2009, ch.7.
[128] A more Humean approach to autonomy might content itself with an idea of psychic unity achieved through steady affective inclinations, which may well be a mere product of impulses we are passively subject to. On such a view, the unity of rational agents would be a matter of their unity *qua* affective creatures.
[129] The term "series" is Korsgaard's (2009: 76). She also uses "heap," likewise implying a lack of integrative structure (ibid.).

on the other, the idea that one's dispositions of reasoning might unify one's mind *qua* rational agent begins to look quite appealing.

The question is, of course, how anything in this imports the need for constitutive norms, or for CI in particular. The very idea of a disposition of reasoning is already the idea of something with internal cohesion: there is a sort of internal unity to one's way of going on in following a rule. Furthermore, a disposition of reasoning is not a momentary thing, acquired in a flash and lost in a moment. This stability of dispositions accounts for a sort of diachronic unity within the agent: a diachronic unity that is a matter of having a reasonably stable way of going on, a way of seeing, and of continuing to see, certain considerations as grounds for certain actions. But nothing in these points about unity favors one rule over another.

So why does Korsgaard think that being a unified agent requires following CI in particular? Her argument employs the notion of *identification* with certain members of one's psychic economy. Korsgaard argues that unified agency is impossible without such identification, and that such identification is impossible without following CI.[130]

3.3.2 Synchronic unity, identification, and weak universality

To see how the argument goes, we need to understand some of Korsgaard's background theory of action. Korsgaard thinks that when we choose an action, what we choose is an act-for-the-sake-of-an-end: an act-for-the-sake-of-an-end is what an action is. And to choose an action is always to choose to enact a *maxim*, a proposed act-for-the-sake-of-an-end (2009: 10-12). But Korsgaard also thinks we are continually passively subject to various desires and impulses, because as well as being rational beings, we are also *animals* with instincts that render certain features of our environment appealing or aversive for us. That is, our instincts render certain features of our environment "incentives" for us (2009: 104-5). Now, our

[130] Though as I argue, the notion of identification is in fact ultimately superfluous in Korsgaard's argument.

animality and our rationality interact in a complex way. The incentives that our instincts make us subject to propose maxims for us: they "prompt" us to consider an action in prospect (2009: 75, 105-7, 116). But where a non-rational animal's instincts directly determine what the animal does in the face of which incentives, rationality introduces a "reflective distance" between your incentives and your response. You are now faced with the question what to "count as a reason," which incentive to act on (2009: 115-6). And settling that question is a task for your rational capacities. What you must do *qua* rational agent is to choose among proposed maxims, thereby choosing to act on the incentive that prompted you to consider that maxim.

This is where the notion of identification comes in for Korsgaard. To choose to enact a specific maxim, Korsgaard thinks, you must "identify with" that maxim. This is because, unless you do so, you must regard the proposed maxim as merely another "force on a par with" your various desires and impulses. Your psyche would be just a field for a battle between various forces, a battle to which you yourself are a mere "passive spectator." There would be no unified self over and above these forces, a self that determines itself to act. The idea that "self-determination requires […] identification with the [maxim] on which you act" (2009: 75) is supposed to explain how one can take command of the battle field of psychic forces and act on some of them.[131] What Korsgaard then argues is that such identification is impossible except through deliberating in accord with CI—that is, through the principle of acting only on maxims whose form is universal.

Why does identification require acting on maxims with universal form, and what is it for maxims to have universal form? Deliberation in general, on Korsgaard's picture, is the business of deciding which

---

[131] Korsgaard sometimes talks about "identifying with" one's *incentives*, and at other times about "identifying with" the "principles of choice on which you act" (both occur at 2009: 75, for example); and it is sometimes unclear whether, by the "principles of choice on which you act" she means one's *maxims* or one's principles of reasoning (that is, CI+HI). I think the most cohesive way to interpret her discussion at pp. 75-6 is as being concerned with the idea of identifying with maxims; or perhaps, with the idea of identifying with incentives *qua* ingredients in maxims (see below). In a different context, Korsgaard also talks about identifying with one's "constitution," which I take to refer to the "constitutional procedure," or CI (2009: 134); but I do not understand what it might be for an agent to "identify with" CI, nor what possible role "identifying with" CI could play in the theory, unless it is just an unhelpful way of saying that the agent necessarily follows CI. For related discussion with regard to "self-legislation," see §3.4 below.

maxim to act on. To do that, she thinks, we have to test the "form" of the maxim.[132] And according to Korsgaard, there are three putative "forms" that maxims might have, corresponding to three putative modes of willing: particular, general, and universal. Particularistic willing would be acting on a maxim that you take to apply "only to the case before you," as having "no implications for any other case" (2009: 72-73). General willing would be a matter of acting on a maxim you take to apply to a "wide range of similar cases" (2009: 73). Universal willing comes in two varieties: in willing a maxim as *absolutely universal*, you take the maxim to apply to "absolutely every case of" the exact same sort. In willing a maxim as *provisionally universal*, you aspire to absolute universality, but acknowledge that you might not have thought of everything that is relevant to the case, and so might need to go back and revise the maxim to give it absolutely universal form (2009: 73). What Korsgaard is trying to argue for is that only universal willing is possible, because only through universal willing can we "identify with" the maxims of our actions. So why exactly can't we will our maxims as particular or general instead?

Puzzlingly, having mentioned the category of general willing, Korsgaard never returns to it: she only argues that particularistic willing is impossible, and claims that this means that the only possible sort of willing is universal willing (2009: 75-6). Despite this oversight, Korsgaard's argument against particularistic willing is of some interest, as it looks to also provide an independent argument against general willing. I will quote the argument in full:

> In order to will particularistically, you [would have to] in each case wholly identify with the incentive of your action. [But] you [could] not identify with the incentive as representative of any sort of type, since if you took it as a representative of a type you would be taking it as universal. For instance, you couldn't say that you decided to act on the inclination of the moment, *because you were so inclined*. Someone who takes "I shall do the things I am inclined to do, simply because I am inclined to do them" as his maxim has adopted a universal principle, not a particular one: he has the principle of treating his inclinations as such as reasons. A truly particularistic will must embrace the incentive in its full particularity: it, in no way that is further describable, is the law of such a will. But this means that particularistic willing eradicates the distinction between a person and the incentives on which he acts. And then there is nothing left here that is the person, the agent […] as distinct from the play of incentives within him. If you have a particularistic will, you are not one person, but a series, a *mere heap*, of unrelated impulses. […] Particularistic willing lacks a subject, a person who is the cause of his actions. So particularistic willing isn't willing at all. (2009: 75-6)

---

[132] For the "testing" model of deliberation, see Korsgaard 2009: ch.3. I return to this below.

The argument purports to be a *reductio* of the idea of particularistic willing. It is somewhat hard to parse, since Korsgaard now claims that a particularistic willer would have to identify with her *incentives*, whereas before the object of "identification" was supposed to be a *maxim*. But perhaps the thought is that an incentive is somehow incorporated into the maxim, as one's ground for the action; so that what we end up choosing, whenever we choose anything, are maxims of the form: "Do act A for the sake of end E, and on the ground of incentive I." This is certainly suggested by Korsgaard's remarks around the middle of the quoted paragraph.[133]

With this amendment to Korsgaard's official picture, the argument in the passage just quoted looks to contain not only a *reductio* of the idea of particularistic willing, but also of the idea of general willing. Korsgaard's idea seems to be that if you so much as think of your incentives as representatives of a type, you are already thinking of your maxim as universal. Or, more carefully: if you so much as think of act A, end E, and incentive I as representatives of a type, you are already thinking of your maxim as universal, because you must be thinking of it as applying whenever the acts, ends, and incentives proposed in a maxim are representatives of the types A, E, or I. But if thinking of the ingredients in a maxim as representatives of a type is all it takes to regard a maxim as universal, then surely you cannot even think of a maxim at all *except* as universal. For in so much as thinking of an act, end or incentive as *falling under a concept*, you regard it as representative of a type. The only exception to this claim would be thinking of the act or end or incentive as falling under a demonstrative concept, such as the concept *this*. But it is hard to see what we could point to and how, to indicate the object we mean with 'this' in considering a putative maxim. The act and the end are still merely in prospect, not yet anything in the world; and neither they nor one's incentive seem to be objects that one could somehow point to through some sort of an act of mental ostension (if the very idea of mental ostension even makes sense).[134]

---

[133] Compare Allison's famous "Incorporation Thesis," (1990: 5-6), which he finds in Kant's *Religion* 6: 24.
[134] Perhaps we could instead give singular names to the acts, ends and incentives in question? But this seems to depend on ostension. In formulating the argument in this paragraph, I drew inspiration not only from the cited

If this is right, then no wonder there is no room for general willing. If you cannot so much as *think* of a maxim except as universal, then surely you cannot choose to enact it except as universal, either. For surely you have to be able to at least entertain a maxim, and so think of it, in order to choose to enact it. So the only possible willing is universal willing: not only is there no room for general willing, but neither is there any room for particularistic willing.

Strikingly, this argument does not even need the notion of "identification." To be sure, Korsgaard's own version of the *reductio* against particularistic willing does seem to employ that notion. Her idea seems to be that the particularistic willer would have to attempt to identify with her maxim, or her incentive, "in its full particularity," not as representative of a type; and that this would really be no different from being moved by the "play of incentives within," one incentive pushing this way, another that. This, in turn, would eradicate the basis for thinking that there is an agent "over and above" the "play of incentives within" at all. But it is hard to see what the idea of "identification" really does even in this argument: it seems to be a mere stand-in for choice. To act on a maxim, one must choose to enact it, and this means that there must be someone, an agent, who does the choosing—someone "over and above" the "play of incentives" that prompt one to consider that maxim. If failure to think of one's incentives or maxims as representative of a type eradicates the distinction between oneself and one's incentives, then in failing to think of one's incentives or maxims as representative of a type, one is not someone "over and above" the "play of incentives." And if thinking of one's incentives or maxims as representative of a type is, on the other hand, enough to make one regard them as universal, then one can only act on them as universal. It is hard to see what the idea of "identification" adds to these thoughts.[135]

---

passage, but also from 2009: 76, fn. 20. Unfortunately I cannot say that I captured the point Korsgaard intends to convey: I am pretty sure I do not quite understand the passage, or the footnote.

[135] Though I cannot argue for this here, the notion of "identification" seems likewise superfluous as used by Harry Frankfurt and others in the tradition of "hierarchical" models of autonomy. One is supposed to be "active" with respect to one's desires, and to be the author of the actions that those desires supposedly prompt, when one "identifies" with those desires by having some type of a second-order attitude towards the desire. The notorious problem with hierarchical models is, of course, to explain how mere addition of higher-order structure is enough to launch us out of a stance of "passivity" to one of "activity" or "authorship" of our actions. The feat has been attempted by diverse characterizations of what "identification" actually amounts to; but then it is those

So what to make of Korsgaard's argument for CI as a constitutive norm of agency? My objection to it is simple. If deliberation is indeed a matter of "testing" maxims for their form, then it does seem that, given the weak criterion of "universality" operative in Korsgaard's *reductio* of particularistic willing, universal willing is the only possible sort of willing. But that weak criterion of universality makes it impossible to see CI as a putative *norm* for willing. If action is always action on a maxim, as Korsgaard claims; and if action on a maxim is impossible without thinking of that maxim; then action on a maxim is *always* action on a maxim that is "universal," in the proposed sense. On the criterion of universal willing that is weak enough to rule out the possibility of general and particular willing, we cannot fail to will universally, if we "will" at all. Of course, there may be failures in execution: we can always fall flat on our faces in attempting to do what we decided to do. But that is not a failure of rationality: it is a failure of dexterity. There is no room for deliberation that ends in action but that does not end as CI says it should—in choosing to enact a maxim with universal form. This means that there is no possibility of following CI while acting in a way that is wrong by its lights. CI lacks rule-internal normativity. Since rule-internal normativity is a prerequisite of authority, CI cannot be a practical norm.

Korsgaard is, then, faced with a dilemma. *Either* she must tighten her criterion of "universality" but admit that, for all that she has said, someone's deliberative principle or principles might be merely general ones; *or* she must admit that CI is not capable of being a practical norm. Both horns of the dilemma are destructive to the proposal that CI is a constitutive norm of agency. The first horn is destructive in challenging the claim that it is constitutive of rational agency to follow CI at all. The second horn is destructive in challenging the claim that, in the sense in which CI might be constitutive of agency, it could be normative.

One might object that failing by CI's lights is still possible on Korsgaard's picture, if all that is required for acting is that one's actions issue from a *disposition* to act on universal maxims, not that one's maxims are actually universal. But the trouble with Korsgaard's argument against particularistic willing

---

characterizations, not the notion of "identification," that do the work in the account. For Frankfurt on identification, see e.g. his 1977 and 1987.

is, precisely, that it relies on a criterion of universality of maxims that is so weak that it makes alternatives to universal willing strictly impossible. If action is always action on a maxim, it is hard to see how one could act on a maxim at all without acting on a maxim that is universal in the weak sense operative in the argument.

So is there available a tightening of the criterion of universality such that we could argue that CI is both normative and constitutive of deliberation as such? The weak interpretation of universality that figures in Korsgaard's *reductio* of particularistic willing certainly looks alien to Kant, whose central ideas about agency Korsgaard takes herself to be elaborating. Furthermore, the weak notion operative in the *reductio* does not even seem to amount to the notion of universality Korsgaard herself first introduced in contrasting universal willing with particular and general willing. According to her initial definition, recall, universal willing comes in two varieties. In willing a maxim "as" *absolutely universal*, you take the maxim to apply to "absolutely every case of" the exact same sort. In willing a maxim "as" *provisionally universal*, you aspire to absolute universality, but acknowledge that you might not have thought of everything that is relevant to the case, and so might need to go back and revise the maxim to give it absolutely universal form (2009: 73). General willing, in turn, is action on a maxim you "take to apply to a wide range of similar cases"; and particularistic willing is action on a maxim you "take to apply only to the case before you" (2009: 72-3). These definitions certainly seem to leave room for the possibility that one might formulate or propose a maxim, and in so doing think of it as having to do with acts, ends, and incentives of certain types, *without* yet thereby willing the maxim *as* "applying" to "absolutely every case of the exact same sort," even provisionally. Universal willing does not seem to be a matter, merely, of thinking of the maxim's ingredients as falling under certain types of act, end, and incentive; but rather, of willing the entire maxim, once formulated, "as" universal, where this is a matter of thinking of the whole maxim as having a universal *status* in some further sense.

But what sense is this? And further, what grounds do we have, on a strengthened criterion for universal willing, for precluding the possibility of particularistic or general willing? Why can I not regard my maxim, already formulated, and my incentive, already grasped as being of a type (say, the prospect of

some forbidden pleasure), as ones on which I shall act *just this once, and never again*, even if the very same circumstances arise? Why can I not regard my maxim and incentive as ones I will act on *generally* but not always, making exceptions when it suits me? I cannot see that anything in Korsgaard's argument against particularistic willing actually speaks to these questions. In acting through a disposition of reasoning, one is not simply a passive observer to one's impulses: one precisely acts *from a principle* that is one's "constitutional procedure," or one of one's "constitutional procedures." It does not seem to make a difference to this point what one's principle is. Even the principle of the wanton who acts on every incentive he can still looks to be a principle—only a maximally permissive one, so far as one's incentives are concerned. (Of course, such a principle might not be capable of being normative, if it rules nothing out.)

However, Korsgaard's argument against particularistic willing is primarily concerned, I take it, with the conditions of *synchronic* unity in agents: with the conditions of being, at the moment of willing, something "over and above" a mere heap of impulses. Korsgaard does have a further argument for a more stringent criterion of universality, based on the conditions of *diachronic* unity in agency. So let us see whether the conditions of diachronic unity fund a conception of universal willing on which CI satisfies the dual conditions of rule-internal normativity and metaphysical necessity required for it to be a constitutive norm. As I argue, the conditions of diachronic unity do not ultimately help in this regard; nor do the conditions of either synchronic or diachronic unity give rise to any other constitutive norm besides CI.

3.3.3 Diachronic unity and strong universality

Korsgaard models her account of the conditions of diachronic unity in individual agency on what she regards as the conditions of collective agency—of acting as one together with others.

Korsgaard thinks collective agency is a matter of deliberation and decision that is "shared" in a quite literal sense. In collective agency, two or more people "deliberate together" in a way that is not just

a matter of individual people deliberating independently, from their own private points of view, and attempting to negotiate with each other so as to come to a compromise acceptable to each party. Deliberation in collective agency is, rather, in some essential sense "shared." In what sense? According to Korsgaard, shared deliberation is a matter of each party acting on *reasons* that are shared, or "public." If I act on a "private" reason, I regard it as having "normative force" only for me, now; and if I regard your reasons as private for you, I regard them as having "normative force" only for you. Insofar as we both regard our own and each other's reasons as private, neither of us takes the other person's reasons into account as good reasons in our own deliberations, though we might regard them as tools or obstacles in relation to our own respective private projects. In contrast, if you and I act on public reasons together, we regard our reasons as good reasons for all concerned: I take your would-be private reasons and treat them as having "normative force" for me as well as for you, and you do the same for me. The result is that the reasons we act on are themselves shared (2009: 191-196). To illustrate the contrast and to connect it to the issue of universalizability, Korsgaard considers an example where you and I both want some object very badly:

> [Suppose] I think I have a reason to shoot you, so that I can get the object. On the private conception of reasons, universalizability commits me to thinking you also have a reason to shoot me, so that you can get the object. I simply acknowledge that fact, and conclude that the two of us are at war. Since I think you really do have a reason to shoot me, I think I'd better try very hard to shoot you first. But on the public conception of reasons, we do not get this result. On the public conception I must take your reasons as my own. So if I am to think I have a reason to shoot you, I must be able to will that you should shoot me. Since presumably I can't will that [consistently with wanting to have the object], I can't think I have a reason to shoot you. (2009: 191-192)

Korsgaard's thought is that acting on private reasons leaves us in a "condition of essential conflict" (2009: 191), unable to form the unified will required for true collective action. I am stuck with the reasons I think are good reasons for me, and you with the reasons you think are good reasons for you. While I can "universalize" my maxim in the sense of thinking that you, similarly situated, would have private reason for exactly the same sort of violence that I take myself to have private reason for now, it remains the case that, so long as our reasons are private ones, we cannot get true collective action into view. In acting on public reasons, in contrast, we transcend this predicament. You and I can literally deliberate together about what to do with regard to the object we both want, and if such deliberation goes well, it will result

in a shared decision. (We might agree, say, to share the use of the object.) In fact, in acting on public reasons, I *cannot* continue to want the object solely for myself; for in taking your reasons "as my own" and acting on them together with you, I thereby will that you should have use of the object too. Likewise for you, if you act on public reasons. In acting on public reasons, we universalize our maxims in the sense that we act only in ways that everyone concerned could simultaneously regard themselves and each other to have good reason to act. This is what happens in true collective action, according to Korsgaard: we form one "unified will" which decides and acts (2009: 189). The maxim we enact is enacted as "public law" (2009: 204).[136]

Now, Korsgaard thinks that acting on public reasons gets us into recognizably "moral territory," for in acting together on public reasons, we both take into account each other's points of view as well as our own (2009: 192). In fact, however, even if we think of collective action in Korsgaard's way, this does not require us to think of the collective agent as anything but a *club*: in taking *your* reasons as mine, I need not also take my neighbor's reasons as mine (or as ours). Moreover, certainly Korsgaard's argument thus far says nothing about why anyone *ought* to partake in collective action of any sort, nor about why agents as such must, metaphysically speaking, be disposed to do so. Finally, for all that Korsgaard has said so far, it looks possible, for instance, to be so egotistic as to consistently regard my private reasons as not only my concern but also as *your* concern, while refusing to recognize you as having any private reasons of your own at all. (Or: while refusing to recognize you as having any private reasons that I should take into account.) If I have this attitude towards my fellows, I regard them as mere tools or

[136] Korsgaard's employment of "reasons"-talk is sometimes confusing. Her official view is that reasons *just are* maxims: in acting on a reason, one simply acts on a maxim (2009: 12-14). I take it, then, that in regarding one's reasons as public, one regards one's maxims as public. On this way of thinking, it looks as though Korsgaard is saying that there are two different possible ways to regard a maxim as universal: as private but having private application to everyone else in similar circumstances, or as public and universal in the sense of being a maxim that we could will *qua* collective agent. Sometimes, however, Korsgaard speaks of "reasons" as though they are reasons *for* adopting or enacting one maxim rather than another, as "considerations with normative force" (e.g. at 2009: 202, 203).

obstacles to my own individual agency, bothersome perhaps but there to do my bidding or else get out of my way.[137]

Nonetheless, we can see that if the diachronic unity of individual agency is to be modeled after Korsgaard's view of collective action; and if at least some degree of diachronic unity is metaphysically necessary for individual agency, as we admitted; then each individual agent must at least be disposed to act on reasons that are "public" with respect to stages of herself other than her present self. And this, Korsgaard thinks, requires each of us to enact only maxims whose form is at least provisionally universal, in the sense that all relevant "stages" of oneself could will action on those maxims, barring discovery of good reasons why not:

> [A]cting is quite literally interacting with yourself. The requirements for unifying your agency internally are the same as the requirements for unifying your agency with that of others. Constituting your own agency is a matter of choosing only those reasons you can share with yourself. That's why you have to will universally, because the reason you act on now, the law you make for yourself *now*, must be one you can will to act on again *later*, come what may, unless you come to see that there's a good reason to change it. (2009: 202-3)

Korsgaard illustrates with reference to Parfit's famous Russian nobleman case.[138] A young socialist, the nobleman wishes to distribute to the peasants the estate he will inherit; but anticipating himself to turn conservative by the time he actually comes upon the inheritance, he signs a legal contract which will make the distribution automatic, and which "can be revoked only with his wife's consent" (Parfit 1984: 327). The nobleman then makes his wife promise not to give in to his later, conservative self's pleadings to revoke the contract: by this time, the wife is to regard the man asking for the promise as effectively dead, and thereby as unable to release her from the promise.

Korsgaard's attitude to the nobleman is that he "fails as an agent": for he fails to will a maxim "that he thinks he can commit himself to acting on again later on, come what may" (Korsgaard 2009: 203). He is at odds with himself, internally divided across time. In fact, Korsgaard thinks, the nobleman is also at odds with himself in the *present* in a way that mars his agency. For try as he might, he cannot *now*

[137] Korsgaard does later argue, on the basis of the conditions of unity in *individual* agency, that in fact we are all committed to collective action with all of humanity, and so to willing our maxims as "public law" for all (see fn. 139 below). But I argue against Korsgaard's views on individual unity below.
[138] Parfit 1984: 327-9.

make a "law" for himself to distribute the estate in the future, unless he regards his maxim of distributing the estate as "public" in the sense of having "normative force for his later self" (2009: 204). And this he fails to do, Korsgaard thinks, because he expects to "change his mind without a reason" and to need his wife's help in heeding his own decision: his mind is a disunified heap, nothing over and above the promptings of momentary "private" reasons, or would-be reasons (2009: 203-204). If making a decision now for the future is to be possible at all, Korsgaard concludes, we must will our maxims as universally binding (if only provisionally) over all stages of ourselves from here on in: as having "normative force" not only over our present selves, but also over our future selves.[139]

What to make of this argument? To be sure, the Russian nobleman changes over time. This may seem like a kind of diachronic disunity. Certainly his anticipated future self has a very different way of looking at things than his present self. More strongly, the nobleman himself supposedly regards any possible future conservative version of himself as no longer an instance of himself at all. If this opinion is correct, then the nobleman is *very* disunified diachronically: "he" is two entirely different people. This latter is surely a possibility that an account of agency should not rule out. Agents do go out of existence; and perhaps this sometimes happens without the agent's body expiring. In such cases, one may well find oneself thinking about what to do about, or perhaps together with, the future inhabitants of one's body. Nonetheless, if the issue we are interested in is one of unity in individual agency, then we must not assume at the outset that the nobleman *is* two different people. What we should assume is just that his anticipated future self is qualitatively very different from his present self, in taking there to be different reasons for different actions in the circumstance; and what we must ask is whether and under what conditions the nobleman's present decisions for his future can unify his agency across time, or indeed in

[139] Korsgaard further glosses this as a matter of "respecting the humanity in yourself"; and she argues that respecting the humanity in yourself depends upon respecting humanity in general (2009: 204). The result is supposed to be that we must will our maxims as universally binding not only over our own selves across time but over everyone—as "public law" for each and every person (2009: 204-206). But I will not go into this further argument here, since I think that Korsgaard goes wrong at an earlier stage.

the present, regardless of those differences. Is Korsgaard right to regard universal willing as a condition of decision for the future? And if so, does this amount to a constitutive norm of agency?

One option, of course, would be to deny that agents as such must be capable of making and acting on decisions for the future at all. Perhaps there can be agents who only make decisions that they carry out even as they make them. But to make sense of agents who perform actions that take any time—and this seems to be most actions available to ordinary agents—we do need some sense of diachronic unity of agency. We need to require at least that an agent initiating an action does not, in every case, stop acting immediately because of disowning her previous decision. Action that takes time is a microcosm of decision for the future: at each moment, it seems, one *might* change one's mind and reverse one's earlier decision. What seems to be required for agency is that one does not always do this—whether because one does not always change one's mind, or because one at least relies on devices such as the nobleman's to carry out one's earlier decisions even as one changes one's mind. The question is whether anything in any of this imports a constitutive norm.

So let us ask again: is Korsgaard right that decision for the future requires universal willing, and that the nobleman cannot make such decisions because he fails to will universally? Korsgaard's interpretation of the nobleman's predicament is odd. Korsgaard claims that the nobleman is unable to regard his maxim of redistributing the estate as "public" in the sense of having "normative force for his later self"; that this is because the nobleman expects to change his mind without any reason; and that this renders him a mere heap (2009: 203-4). But why think that, even if the nobleman does expect to change his mind for no reason, this expectation is an accurate reflection of reality in some way that renders him, in fact, a mere heap? More importantly, surely the nobleman *does*, now, regard his decision as having "normative force for his later self." He thinks it is a good decision for the future. His problem is just that he anticipates his later self to disagree, and therefore to reverse the decision. (Of course, it also seems that *part* of the nobleman's problem is a puzzling attraction to complicated arrangements: why include, in the legal document he signs, a provision to the effect that the document *can* be revoked at all, even with his wife's consent?)

Crucially, then, it does not seem that the nobleman needs to expect himself to simply reverse course "without a reason," contrary to what Korsgaard claims. The nobleman can fully anticipate his future self to have reasons to resist redistributing the estate—by his present lights bad reasons, but reasons all the same, and by his future lights good ones. His future self's reasons would be, simply, all the reasons why a different, conservatively-minded nobleman might resist a suggestion, now, to redistribute his land to the workers. "They don't deserve it," he might say; or "What did my father work for all his life if I give it all away now?" or "Socialism is a crock." Neither the future nor the past nobleman's reasons and the way they hang together individually seem like a mere "heap": they are simply very different from each other.

So if there is some reason why we must think that the nobleman fails to decide for the future now, it is not that he must regard his future self as changing his mind for no reason, in some way that warrants our regarding his present self as a "mere heap." To the contrary, surely he regards his future self as changing his mind for *bad* reasons, and correspondingly, he regards his present decision as a good one for the future—as "having normative force for his later self." That he regards his future reasons as bad ones and his present reasons as good ones is plausibly precisely *why* he goes to such lengths to ensure that the anticipated change of mind will remain inefficacious. (If it was all the same to him, his determination to have his present decision enforced instead of his anticipated future one would be strangely arbitrary.)

More strongly, it does seem that our nobleman *can* in fact decide for the future. If he had absolutely no way of enforcing his present decision while fully expecting to revoke it later, things might be different: it would be rather like intending to φ while fully believing that one will not φ—a feat that looks impossible.[140] But as it is, the nobleman has his ways. Distrusting his resolve, he makes external arrangements to carry out his present will. Something that *is* wrong with the nobleman is his lack of resolve, and his bad faith about what he can do of his own resources to affect the future self he becomes.

---

[140] On this point, see Kavka 1983.

But this does not, apparently, affect his ability to decide for the future, given that he can enforce his present decisions for the future through legal means and through his wife's help.

In Korsgaard's terms, then, it seems that the nobleman can regard his maxim as "public" at least in the sense of "having normative force for his later self" (2009: 204). It is just that he thinks his later self will disagree about the authority or goodness of the prior decision.[141] Notably, however, the sense of 'public' that is present both here and in Korsgaard's diagnosis of the nobleman's predicament is much thinner than the sense of 'public' one would have expected based on Korsgaard's discussion of collective agency. In unilaterally regarding his present reasons and decisions as good ones, and as to be enforced despite his later self's anticipated disagreement, the nobleman seems more analogous to the egotist in our earlier example than to someone who seeks to "unify his will" with others. The nobleman regards his later self as a bothersome obstacle to his present decision, or as someone whose hand must be forced. It is as if I regarded my reasons as your concern, but your reasons as no-one's concern. If this is a case of "universal" willing, or of willing a "public law," it is quite disanalogous to the case in which, in deciding which maxim to enact, I must consider whether me and my fellows could or would all enact a maxim together after "shared deliberation."

So even though the nobleman can decide for the future, he need not will universally in the robust sense involved in Korsgaardian "shared deliberation" with others. Nonetheless, we can ask: is the sense of "universal" willing required for decisions for the future, even the nobleman's, still robust enough to import a constitutive norm? The crucial issue here is whether the conditions of willing one's maxim as having "normative force" universally over one's future selves as well as one's present self in any way restrict the range of maxims that one may will as having such force. If *any* maxim can be willed as having

[141] Something that may be at work in Korsgaard's argument is her view that "normativity" is not only authority but also a "psychological force" (2009: 2-3). But that view of normativity is obviously very controversial, and simply posits the kind of tight link between normative facts and psychological facts that the constitutivist is trying to defend. It may be true that the nobleman's future self will not experience his present decision as having any psychological hold on him. But that does not mean that he could not or should not do so. In any case, the present nobleman can certainly *think* that his future self should act on his present decision.

"normative force" for the future, then the conditions of decision for the future cannot help us distinguish between good deliberation and bad, good maxims and bad.[142]

What effected the restriction on the maxims that *collective* agents can permissibly will was that each member of the collective agent must regard the others' private reasons "as their own." This restricted the permissible "public maxims" of the collective agent to ones that cohere with each individual member's private reasons. Of course, the collective agent might go wrong in its deliberations, and end up disproportionately disadvantaging one of its members because of a miscalculation. But the point is that this would be bad or defective action by the lights of the strong constitutive requirement of universalization, on which we each take each other's reasons as our own, that governs collective action as such.

However, Korsgaard explicitly distances herself from interpreting decision for the future along analogous lines. She says we need not regard the Russian nobleman's only way to unify his will as forging a "dreary" compromise between the private reasons of his various stages—though this would certainly be one way for him to unify his will (2009: 203). It seems to me that Korsgaard is right to retreat from the idea that what is required is a compromise. Surely we *can* unilaterally decide for the future without so much as consulting the private reasons of our future selves. In any case, it can be hard to discern one's future "private" reasons; and even in cases where I think I know my future private reasons—the things that will seem to me worth acting on come weekend, say—I can unilaterally decide now to ignore them. (A different question is whether the decision will be effective come weekend; more on this below.) Often it is precisely *because* one thinks of one's present reasons as the good ones, as having "normative force" for the future, that one makes a present decision for the future that ignores or discounts one's anticipated future "private" reasons. The question is whether anything in this weaker

[142] One might object that it is not even a requirement for deciding for the future that one must regard one's decision, or one's reasons, as "having normative force." Perhaps it is enough that one simply resolves to do something, no matter how horrible one regards one's present decision or reasons for it as being. But as I argue, even if all decision, including decision for the future, requires one to regard one's reasons and decisions as in some way good, this does not import any constitutive norms.

condition of decision for the future, that one must regard one's (unilateral) present decision as having "normative force" in the future as well as now, imports any actual restriction on what one can will for the future compatibly with fulfilling this condition.[143]

No such restriction is introduced. Consider a different nobleman, a conservative young man who expects himself to soften in his old age and to become "a sentimental old socialist fool," as he regards his expected future self. Here, too, the conservative young nobleman could make a unilateral decision for the future, regarding his present (conservative) reasons as good ones and his future (socialist) reasons as bad ones. In so doing, he would regard his present decision as a good one, and as having "normative force" for the future: he thinks it ought to be carried out. Distrusting his future self to carry it out of his own accord, he makes legal arrangements instead. His present maxim—"Sign a document now to keep the land later," or something to that effect—is diametrically opposed to the present maxim of his socialist youth counterpart. Yet they can both will their respective maxims for their respective futures. And it seems that we could substitute anything for 'socialist' and 'conservative' and get essentially the same result. Expecting to turn dishonest later, a young honest politician can decide now to make arrangements for her later self to be held to a standard of accountability and transparency; and a present corrupt politician might take steps now to try to stop her anticipated future self with a guilty conscience from giving herself in.

We can put the point more generally, abstracting away from the circumstance of someone seeking to circumvent her own future action on her future "private" reasons. Many ordinary decisions for the future do not involve one in maneuvering to put in place external restrictions on one's future self's ability to revoke one's present decision. (Struggling with addiction, or less severely, against temptation, is of

---

[143] One restriction that does seem to be in place is that one's maxim must concern a future circumstance one regards as possible. This is just a version of a general restriction on what one may decide: one's decisions must be for actions that one regards as possible, in circumstances that one regards as possible. But while that restricts the range of maxims one can will, the requirement does not seem violable: one cannot so much as attempt to decide for the future while flouting this requirement. So it cannot be a norm for decisions for the future. Nor does the requirement stem from the criterion of "publicity" or "normative force" under consideration.

course an ordinary case that is more akin to the nobleman's.) Nonetheless, even in such cases, the thought that one must regard one's decision for the future as binding for the future does not, in and of itself, delimit the range of reasons one can now take to be good, or the range of decisions one can now make for the future. The putative norm of deliberation "Take your reasons to be good ones, and take your decisions for the future to be binding for the future," in and of itself rules out *no* reasons or decisions as ones that someone might take to be good or binding. And if it rules nothing out, it cannot be a practical norm governing deliberation and decision. The result is that even if it is a constitutive fact about decision for the future that one must take one's present decision, as one makes it, to be binding for the future, and one's present reasons to be good ones, this constitutive fact about decision for the future is not thereby a constitutive *norm*.

What *does* delimit the range of reasons and decisions one takes to be good, and so the decisions one actually makes, are one's specific practical dispositions to take certain considerations, *p*, as good reasons for certain actions, φ, in certain ways. In fact, *some* specific way in which one takes *p* to be a good reason to φ is required for one's taking *p* to be a good reason to φ to even make sense. One cannot simply take anything, say, the fact that the grass is green, as a good reason for just any decision, say, to become a psychiatrist, unless we can tell some back story to make it intelligible *how* that sort of thing, the color of grass, could be taken by someone to be a reason for that sort of thing, becoming a psychiatrist. We cannot magic the requisite intelligibility into existence simply by saying that one takes one's reason to be a good one.[144] It does not help if one says "It's not just that *p* is a good reason for me *now* to φ; it's also a good reason for me to φ *in the future*." There still has to be some *way* in which one takes considerations *p* to be good reasons to φ. But as with the socialist and the conservative, the honest and the corrupt, it seems that the ways in which one takes specific considerations to be good reasons for specific actions or decisions is a contingent fact about oneself. It is, in any case, not something that flows from the conditions of taking one's reasons and decisions as normative for the future, as such. If this is right, then

[144] Cf. Setiya 2010: 97; and fn.124 at p.98 above.

it is enough for decision for the future that one takes one's reasons and one's decision to be good ones *in some way or other*, however contingent one's specific dispositions to do so might be. Nothing in the conditions of decision for the future imports a metaphysical "must" of the sort we could hope to derive a constitutive norm from.

One might object as follows. Even if the requirement of regarding one's decisions for the future as having "normative force" for the future does not place any restrictions on what one may decide, nonetheless the conditions of diachronic unity in agency do place restrictions on what can count as *acting* on a decision, once made. It still matters, for whether one counts as an agent, whether one can follow through with one's decisions. Diachronic unity in agency requires that one actually carry out at least some of one's decisions. And to follow through on a prior decision, one must at least regard one's *past* "private" reasons as one's own, now. So there is at least this requirement of diachronic unity on acting on prior decisions: one must be disposed to carry out one's past decisions, whatever they may be. One must be disposed to heed something like the rule "Carry out your past decisions, whatever they are!"

In response, while such a disposition looks *possible*, it is not necessary for carrying out past decisions. Instead, to carry out one's past decisions, it is enough that one *either* has put in place constraints on one's future self akin to those in the nobleman's case; or, as in ordinary cases, one continues to agree with one's past decisions, and sees no reason (or at least no good enough reason) to modify them as the time comes for action. An agent whose practical dispositions are relatively stable over time is *thereby* of one mind with her past self, in such a way that in acting on her past decision, she regards that decision as hers, now. She stands by it, or "owns" it, as we might say. Such a one need not be disposed to regard her past reasons and decisions to φ as giving her, on their own, any reason to φ now. Her reasons to φ now, if she φ-s, are just her reasons to φ, not the fact that she made a past decision to φ. So the mere fact that one used to think that φ-ing is a good idea does not compel one to think that it is

thereby a good idea now, even defeasibly.[145] Nonetheless, action on a past decision is possible, *qua* action on a decision one remakes in acting on it.[146]

Of course, there are no guarantees, in any given case, that one's practical dispositions will not change, or that one's past decision will not be revoked by one's present or future self. That is a risk one runs in deciding for one's own future. What Parfit's nobleman decides to do in the face of the risk is to circumvent it by forcing his own future hand. That is, as we saw, possible: it is one way of enforcing your decisions. But of course that course of action does deprive one's *future* self of a measure of autonomy with regard to an important decision one could have otherwise taken. So are we stuck between two bad options—on the one hand, having to risk that one's decision will not come to anything, and on the other, making sure that one's decisions do come to fruition but only by constraining one's future autonomy?

I think we are indeed stuck between these two options; but it is easy to exaggerate the badness, as such, of either. It is only the very extreme cases of either that look threatening for one's remaining the same agent through time. And insofar as such extreme cases are possible, as they look to be, the threats they pose to agency are *real* threats, ones that real agents might have to grapple with; they are not mere philosophical problems that a neat theory of autonomous agency should try to sanitize away. At any rate, it is hard to see how the choice between the two options, or something very like them, could be avoided by appeal to a disposition to regard one's past decisions as giving one reason to carry them out, simply because they are past. If that disposition is a disposition to view one's past decisions as giving one merely *defeasible* reasons to carry them out now, then there is still the risk that those reasons are defeated, in one's deliberations, by reasons that one's present self finds more compelling. If, on the other hand, the disposition is a disposition to view one's past decisions as giving one *decisive* reasons to carry them out, then it makes one's present self hostage to past mistakes, depriving her of the ability to reassess her

---

[145] Contrary to Velleman's proposal in his 1997: 237-9.

[146] This does not mean that making decisions for the future is pointless. Often enough, deliberating in advance for action anticipated later does have a point. But my present point is just that, in φ-ing for the reasons that, according to one's prior deliberation, are good ones to φ, one can still regard, in the present, those reasons as good ones for the decision, and make the past decision also one's present one.

reasons and decisions. What is gained in ensuring that one never needs recourse to external constraint is surely lost in viewing agency as such a psychologically rigid affair.[147] It is hard to see why we should think that such a strong disposition is a condition of autonomous agency, or of the diachronic unity required for it.

Why might someone be tempted by the idea that some disposition to act on one's past decisions, simply because they are past, is required for diachronic autonomy? One reason might be a tendency to construe diachronic autonomy as autonomy across a gap in time; as if the crucial issue was securing control of one's future actions by one's past or present self. But we need not, and I think should not, construe diachronic autonomy in this way. Rather, diachronic autonomy, and the diachronic unity required for it, is, or at any rate can be, a matter of one's unity and autonomy *through* time: remaining an agent, stably enough disposed, long enough to act.

3.3.4 Taking stock

I argued, in §3.2, that the general notion of autonomy as self-determination is compatible with the contingency of the practical dispositions through which each agent exercises her autonomy. At the beginning of §3.3, I asked whether this conclusion can be overturned by the conditions of what it is to be a *self* who determines herself to act, an agent "over and above" her various impulses, with a measure of both synchronic and diachronic unity. §3.3.1 suggested, tentatively, that it is enough for both synchronic and diachronic unity that one's dispositions of reasoning are relatively stable through time; and that this is consistent with thinking that it is a contingent matter, from agent to agent, exactly what one's dispositions are. The conditions of unity import no constitutive norms.

---

[147] There is much more to say in this area. In particular, some compulsive dispositions seem to be relatively stable dispositions to do things for reasons, at least in some sense—say, to wash one's hands because one touched something—and such dispositions seem to be psychological threats to autonomy. What distinguishes them from dispositions of reasoning through which one acts autonomously? There may not always be clear-cut distinctions here. But one distinguishing mark might be precisely the rigidity of compulsive dispositions, and their unresponsiveness to reassessments of one's reasons. Unfortunately I cannot delve into this topic here.

Our discussion, in §3.3.2 and §3.3.3, of the conditions of synchronic and diachronic unity has reinforced that conclusion. §3.3.2 considered whether Korsgaard's argument against particularistic willing, and thereby in favor of universal willing, speaks against the conclusion. I argued that the weak sense of "universality" operative in that argument may mean that agents must will "universally," but this sense of universality is too weak to furnish a practical norm. §3.3.3 considered the conditions of diachronic unity and of decision for the future. Korsgaard's work on "public" reasons does provide a stronger notion of universality that does furnish putative practical norms. But I argued that the only sense of "universality" that is plausibly a condition of diachronic unity and decision for the future—namely the condition that one regard one's decision for the future as a good one for the future—fails to furnish a practical norm. Finally, nothing in the conditions of carrying out one's past decisions imports constitutive norms, either. So our conclusion stands: it is enough that one's practical dispositions, whatever they are, remain stable long enough for one to act. There is no putative practical norm implicit in the conditions of agential unity.

As the discussion has progressed, we have inched further and further away from Korsgaard's model of deliberation as "testing" one's maxims for their form, and back towards the idea of a disposition of reasoning as a disposition to make certain considerations *p* one's reasons for certain actions φ in certain circumstances. This latter idea may look to better fit an alternative, "weighing" model of deliberation, a model that Korsgaard rejects. One might then object that we have begged some question against her.

However, much of Korsgaard's own discussion of "public reasons" proceeds, precisely, in terms of reasons for action, not in terms of maxims that one tests. In any case, Korsgaard does not actually argue for the "testing" model of deliberation, except by saying that it "fits" with the idea that practical norms concern the "form" of one's reasoning, not the contents of one's "premises" in reasoning; and by promising that her arguments for the need to follow CI, which is a principle for testing maxims, will vindicate the "testing" model (2009: 52). But of course Korsgaard's first argument, against particularistic willing, is itself premised on the "testing" model; why else should it have looked even remotely plausible that the only possible "forms" of deliberation to consider are particularistic, general, and universal? It is

not a strange or outlandish thought that one's deliberation might have a "form" even on a "weighing" model. The form of one's practical reasoning is simply a matter of the way in which one relates certain "premises" or reasons to certain actions one then undertakes; just as the "form" of one's theoretical reasoning is a matter of the way in which one relates certain premises to certain conclusions.[148]

Having argued against the need to follow CI to secure unity in agency, then, we need not think that our only options are to retreat to a model of deliberation as either "particularistic" or "general." Our options for the "forms" that deliberation can take are wide open. For instance, as I explained in chapter 2, the various character virtues and vices are plausibly dispositions of practical reasoning. Our character concepts seem to capture determinate ways of seeing considerations as reasons for action, and of acting on them. The generous or benevolent person sees another's financial plight as a reason to help. The miserly or mean person sees the person in plight as a likely annoyance and to be avoided; and the positively malevolent person might see the situation as one to relish, and might even plan a scam of some sort to take maximal advantage of the person's plight. Each has a different *way* of responding even to the *same* situation, a different way of seeing features of the situation as reasons for action. We can now add that, in standing for dispositions of practical reasoning, our character concepts signal, in part, that there is a characteristic sort of *unity* to the way that each person goes on. In having a character trait, one has both an internally cohesive way of viewing situations and acting in them, and the cross-temporal unity that the relative stability of a character disposition brings.

I am by no means the first person to suggest that character traits might be dispositions of acting for reasons.[149] All I take myself to be doing here is putting that idea into its proper place in an account of autonomous agency, and connecting it to the assessment of constitutivism. If acting from a character disposition constitutes an intelligible determinate form of rational agency, then there is no specific disposition of reasoning that each rational agent must have to be an agent. This is because there is no

---

[148] Chapter 2 goes into more detail about what it means for one's practical reasoning, or one's practical reasons, to be of a certain "form."

[149] Compare e.g. Aristotle, *Nicomachean Ethics* II 1105b19-1107a6, McDowell 1979, Hursthouse 1999: ch.6, Setiya 2007.

character disposition that each rational agent just as such must have. Accordingly, self-determined action by unified agents can be action through contingent determinate forms of agency.

A last objection to my conclusions concerning the conditions of unified agency charges that, since all agents must tend to be unified to at least some extent, all agents must at least be disposed to be *coherent*, or to follow some rule such as "Unify yourself!", whatever other practical dispositions they might have. My response is three-fold, and necessarily brief.

First, there is a distinction between dispositions that in fact serve to unify you, if you have them, and dispositions that are just dispositions towards unity, as such. Even if all agents necessarily exhibit a measure of psychic unity or coherence, this might be just the net effect of their other, fairly stable dispositions, not the result of a disposition towards unity or coherence as such (Kolodny 2008: 439). Second, a disposition towards unity or coherence as such would be, in Niko Kolodny's words, a strangely "randomizing" disposition to have, since there are so many different ways, in principle, to achieve unity or coherence (2008: 447). If there are other, less "randomizing" ways to achieve the unity or coherence requisite for autonomous agency, then a disposition towards unity or coherence as such is not necessary.[150] Finally, third, we should not overestimate the degree of coherence required for being a unified agent in the sense of being a "self," "over and above" one's various impulses, capable of acting for reasons. It is surely possible to be a somewhat conflicted person and still to act for reasons. One might see some merit in reasons of greed as well as in reasons of avoiding a guilty conscience; and one might have conflicting dispositions to act on each. Still, in acting on either type of reason, one acts for reasons. To the extent that being an autonomous agent is necessary for acting for reasons, the unity required for autonomous agency cannot require absence of such conflicts. But then neither is there any reason, in the

[150] Of course, neither does a disposition towards coherence as such seem to be a disposition of practical reasoning, enjoining certain actions in certain circumstances; it is "wide-scope," not "narrow-scope." If it is not a disposition of practical reasoning, then it cannot help us argue for a view about good practical reasoning and reasons via the constitutivist argument schema. Insofar as good practical reasoning and reasons is our topic, the disposition towards pure coherence as such does not address it.

nature of autonomous agency, to posit a disposition towards coherence as such—a disposition that, if perfectly manifested, would eliminate such conflicts.

The final two sections consider the concepts of *self-legislation* (§3.4), *self-knowledge*, and *self-understanding* (§3.5). If, as I argue, nothing in these concepts imports constitutive norms of agency, then our conclusion stands. The central notions of autonomous agency can be accounted for compatibly with the contingency of the practical dispositions through which autonomous agents act.

3.4 SELF-LEGISLATION

Korsgaard claims that autonomous agents self-legislate, in the sense of choosing "the principles of [their] own causality" (2009: 110). But the idea of self-legislation is ambiguous between two different "laws" or "principles" we might be thought to be in the business of choosing. The first type of "law" is a principle of reasoning, the sort of thing that connects your premises to your conclusions—or, as on Korsgaard's model, tests your maxims. The second type of "law" is a maxim, the "principle" of this particular action. Neither type of "law" helps the constitutivist make use of the idea of self-legislation to yield constitutive norms. Here is why.

The idea of choosing your own principles of reasoning is either incoherent or cannot help the constitutivist. Suppose choice is an exercise of rational agency. And suppose, as we already argued, that rational agency is always exercised in accord with some principle of reasoning or other: there is always some determinate form of reasoning, or of taking things to be reasons, in play. Under these suppositions, it follows that we can never choose our own principles of reasoning, all the way down. The idea of self-legislation all the way down is incoherent. Now, you might try denying the conclusion. Then one or the other supposition must be false. If the first supposition is false, so that the "choice" in question is not an exercise of rational agency, then self-legislation could not belong to rational agency. In that case, the idea of self-legislation could not help the constitutivist argue from the nature of rational agency to practical

norms. If the second supposition is false, so that we can exercise rational agency without following any principles of reasoning at all, then the nature of rational agency *ipso facto* does not have implicit in it any putative practical norms. So again, the idea of self-legislation could not help the constitutivist argue from the nature of rational agency to practical norms.

Korsgaard herself shies away from saying that we choose CI (and HI) as our principle(s) of reasoning (2009: 131). She probably sees that that would defeat her constitutivist ambitions. What she says instead is that we choose the maxims of our actions, which she calls our "laws" or "principles" in a further sense. So do the conditions of self-legislating our maxims import constitutive norms? If we choose actions, and actions are acts-for-the-sake-of-ends, then it is probably harmless to say that in choosing actions we also choose to enact a "maxim," a proposed act-for-the-sake-of-an-end. After all, we can entertain, and wonder whether to perform, actions in prospect. But §3.3 precisely argued that, whether we choose actions through "testing" or through "weighing," nothing in the idea of so doing imports constitutive norms.

In sum, then, the idea of self-legislation cannot topple the claim that autonomous action can proceed through contingent dispositions of practical reasoning. What about self-knowledge and self-understanding?

3.5 SELF-KNOWLEDGE AND VELLEMAN'S SELF-UNDERSTANDING

3.5.1 Spontaneous self-knowledge

Following Elizabeth Anscombe's (1957) way of distinguishing between intentional actions and mere behaviors, David Velleman locates the difference between autonomous actions and mere behaviors in the relation between our behavior and our "forethought about it" (2009: 130-1). Consider two ways in which we can have thoughts about our future behaviors to which our future behaviors conform. One is mere

prediction: the thought "I am going to be sick" may be true of what I am going to do in the next five minutes or so, but its truth does not depend upon my thinking it. (The sense of 'do' here is, of course, a thin one that does not itself yet mark the distinction between action and mere behavior that Velleman is trying to elucidate.) The other is decision: "I am going to take a walk" may be a true thought about what I am going to do in the near future, such that its truth *does* depend upon my thinking it. In the case of autonomous action, "our thinking it makes it so" (2009: 130). In particular, Velleman thinks that the dependence of action on thought is causal dependence: when one's thought that one is going to take a walk is what actually causes one to take a walk, then that thought amounts to knowledge of one's own actions that is "the cause of what it understands" (Velleman 2009: 130, 1996: 194, fn.54; Anscombe 1957: §48).[151] We can sum up these thoughts in the claim that a type of *spontaneous self-knowledge*, a knowledge of one's own actions from which those actions actually spring, is essential to agency.

There are difficulties with this idea, to be sure. Since my thought that I am going to be sick might actually cause me to be sick *via* some strange route, not all beliefs about the future whose truth in some way "depends" upon them make the behaviors that depend upon them autonomous actions.[152] Moreover, in autonomous action, there also seems to be a dependence of *present* action on *present* thought that one is doing it.[153] If one sincerely denies that one is φ-ing, or if one's knowledge that one is φ-ing is extraneous to one's actually φ-ing ["Look! I'm bleeding!"], then φ-ing seems to be something that is merely happening to one in exactly the way that an account of autonomy that respects our starting intuitions about autonomy (p.88) ought to rule out. And it is a good question how we can adapt to the case of present action the idea that the dependence between thought and action is *causal* dependence, without denying that one's thought about what one *is* doing is *knowledge* already as one thinks it. Nonetheless, something in the vicinity of Anscombe's and Velleman's idea has come to seem at least a necessary condition of agency to an increasing number of philosophers of action since Anscombe.

---

[151] Both Velleman and Anscombe attribute the phrase "the cause of what it understands" to Aquinas, *Summa Theologica*, Ia IIae, Q3, art. 5, obj. 1.
[152] This is a version of Donald Davidson's famous "deviant causal chain" worry; see Davidson 1973a: 78-9.
[153] Velleman does formulate his idea about autonomy in terms of present action in his 1996: 194.

I will not defend the centrality of some such idea for autonomous action here. I will merely assume that there is something right in the thought that it is a constitutive fact about action that it necessarily involves one's correct awareness, however implicit, of what one is doing, as one is doing it. Further, I will assume the stronger claim that it is a constitutive fact about acting *for a reason* that, in doing φ because *p*, one has some roughly correct awareness, however implicit, that one is φ-ing *because p*. (One could, perhaps with some effort, correctly explain what one's reasons are in response to questioning.) Lastly, I will assume that the self-knowledge that such implicit awarenesses embody is indeed spontaneous in some way: it is essentially involved in the generation of action, not just an extraneous but necessary accompaniment to acting for a reason as it is happening.[154] Our question is: does something in these putative constitutive facts about autonomous agency also yield a practical norm? And if not, is there some strengthened version of them that is true about agency and does yield a practical norm?

We might suggest that if agency essentially involves spontaneous self-knowledge, then agency as such has the constitutive "aim" of spontaneous self-knowledge. One's practical reasoning could then be judged good or bad in relation to how well it achieves that aim: the rules of reasoning following which helps one to achieve the aim are good ones, and the rules of reasoning following which frustrate the aim are bad ones.

But this avenue does not look very hopeful. If spontaneous self-knowledge is essential to any case of acting for a reason, then one could not act for a reason at all without having spontaneous self-knowledge. One could of course *fail to act*: One's practical reasoning might be interrupted or otherwise inefficacious in actually issuing in action; or one might start to act but be interrupted in the course of the action itself. In such cases, one would fall short of spontaneous *knowledge* that one is φ-ing because *p*: for

---

[154] Pamela Hieronymi expressed to me, in conversation, a view that denies this last assumption: roughly, while there is reason to think that we generally know what we are doing and why, this knowledge is extraneous to action itself, and can be explained as a well-founded prediction based on one's knowledge of one's own mental states, including decisions, and of what they are likely to make one do. As Hieronymi explained to me, the same view might also deny that the extraneous self-knowledge is a *necessary* accompaniment to action: it may be merely a general tendency that action is accompanied by self-knowledge.

one would not be φ-ing.[155] But these facts do not help us distinguish between good practical reasoning and bad, good reasons for action and bad. *All* practical reasoning just as such is reasoning that issues in action; *all* action for a reason just as such is action for a reason. *Any* practical reasoning, no matter what one's rule of reasoning, might be interrupted, and *any* action, no matter what one's reasons for it, might be interrupted.[156] The idea that good practical reasoning and reasons for action are ones that help us attain spontaneous self-knowledge is thus far normatively empty. It rules absolutely no types of practical reasoning "in" as good ones, and no types of practical reasoning "out" as bad ones.[157]

One might object that perhaps the constitutive "aim" of spontaneous self-knowledge is the aim of *maximal* self-knowledge, not just in the sense that one should fully know what one is doing and why as one is doing it, but rather in the sense that one should maximize occasions for such knowledge. This "aim" would amount to a strengthening of the assumptions about self-knowledge I accepted above. It would favor risk-averse types of reasoning, and rule out, as frustrating to the aim, types of practical reasoning that lead one to try risky courses of action. And it would give us reason to perform as many very small and safe actions as we can. It would thus not be normatively empty. But it is hard to see why we should think that agency as such aims at such maximization. Evidently, risky types of practical reasoning are at least possible. The courageous soldier might be led, by reasons of courage, to try a course of action with only slim chances of success. And not only does this seem possible: there does not seem to be anything defective about it, *qua* a case of practical reasoning. If there is some reason why maximizing occasions for self-knowledge is a constitutive aim of agency, I cannot see what it is.

---

[155] One might of course be doing some part of φ-ing, ψ-ing.

[156] At least, this is so for actions that are neither instantaneous nor complete in every instance of one's doing them. But in the case of instantaneous actions, if one acts for a reason at all, then one cannot fail in any way to have the spontaneous self-knowledge we are supposing is necessary for acting for a reason. So cases of instantaneous action are not objections to the anti-constitutivist point I am making here.

[157] One might object that a constitutivist should therefore deny the assumption that the self-knowledge is spontaneous, or in some other way necessary to action: this would allow her to hold that it is at least possible for a case of acting for a reason to fail to achieve self-knowledge. But this would be to admit that self-knowledge is both extraneous and contingent to acting for a reason as such, and so not its constitutive "aim."

3.5.2 Velleman's self-understanding

Velleman suggests a different strengthening of our assumptions about self-knowledge. According to Velleman, autonomous agency as such aims at *self-understanding* in the sense that it aims at what it "makes sense" for one to do "in the light of [one's] circumstances, attitudes, and attributes" (2009: 132). One acts for good reasons if and only if, in doing so, one does what would in fact "make sense" in this way; and for bad reasons if and only if what one does fails to "make sense," or does not "make sense" as well as some other course of action would have. To act is just to seek self-understanding, to seek to "make sense"; and practical reasoning is a matter of attempting to find the action that is most "intelligible" for the agent, where the agent's various characteristics and motives are "clues" to what the most "intelligible" action is (2009: 133). (We might construe such practical reasoning as following the rule "Be intelligible!")

What is the relevant notion of "making sense" or "intelligibility"? Clearly it cannot be just a matter of intelligibility *plain and simple*. No case of acting for a reason could fail by this standard, for surely every case of acting for a reason must be intelligible as one.[158] Nor is the relevant notion of "making sense" normative, Velleman says: it is "not about what [one] ought to do" (2009: 13). Instead, Velleman explains, "making sense" is a matter of acting "in character," where this is a matter of your actions' being intelligible in the causal-explanatory terms of folk-psychology as what such a one as yourself would do in the situation you are in. In acting, one enacts a "self-conception" which includes "everything that the agent knows or thinks about himself," a "description under which your actions and reactions make sense to you in causal-explanatory terms" (2009: 13-14 fn.6, 16 fn.8). Such a self-conception can be false, in which case enacting it yields "self-misunderstanding" and frustrates the aim of

[158] This is not, of course, to say that there are no constraints on what *is* intelligible, plain and simple. I argued above, at §3.2 and at the end of §3.3.3, that there are at least these constraints: the way in which one's action relates to one's reasons for that action must fall into some publicly intelligible pattern, a "rule" loosely construed; and the idea that one's reasons and actions relate via one's conception of one's reasons as *good* ones for the action is not, in and of itself, sufficient to count as such an intelligible pattern.

agency (2009: 26). But to the degree that one's self-conception is true, and one acts in accord with that self-conception, one acts as one should, and for good reasons:

> The understanding that must be possible, if an action is to make sense coming from [this agent], is a folk-psychological understanding that traces the action to its causes in the motives, traits, and other dispositions of the [agent]. […] [In acting, one enacts one's] idea of how it would be understandable for [one] to manifest [one's] thoughts and feelings under the circumstances. […] I believe that the process of […] self-enactment constitutes practical reasoning, the process of choosing an action on the basis of reasons. Why do I think that the self-enactor chooses his action? Because it is his idea, which he puts into action in preference to other ideas that he might have enacted, if this one hadn't made more sense. Why do I think that he chooses for reasons? Because he chooses his action in light of a *rationale* for it, which consists in considerations in light of which the action makes sense. (2009: 13, 18)

Velleman is careful to point out that the self-conception one enacts can be quite subtle: it need not be just a stereotype (2009: 13, fn.6). Nor need the "considerations" that an agent would cite as his reasons, if we asked him why he is doing what he is doing, reflect all the features of the self-conception that constitutes the agent's rationale for the action. Instead, what the agent cites are a few selected considerations that seem to him especially relevant "to understanding his action," or to "directing the questioner's gaze to an angle from which self, situation, and action fall together into a comprehensible *Gestalt*" (2009: 19). Indeed, Velleman clarifies, the *content* of one's reasons for action need not explicitly mention aspects of one's self-conception *at all*: the self-conception that makes sense of our actions for us is in the "background" of our reasoning, not as a target of attention but rather as *structuring* one's reactions to the outward features of one's situation, features which do figure as the content of one's reasons (2009: 19-25).[159]

Why think that autonomous agency does aim at self-understanding in Velleman's sense? Velleman's basic idea is that an account of agency must understand how agents can play a causal role in the generation of their actions, not just how various motives and psychic characteristics of the agent cause behaviors. And he thinks that, to understand this, we must understand the agent himself as essentially embodied in a motive with its own distinctive influence: an inclination or "drive" towards self-understanding. That is, functionally speaking, just what an agent is (1992, 1996: 196; 2009: 17, 133). This

[159] Compare the Sophisticated Humean picture I discuss in chapter 2, §2.3.

allows us to conceive of the agent himself as playing a causal role in altering the balance of his other motives, thus playing an active role in the actions that ensue. An agent's self-conception engages his inclination towards self-understanding, thereby moving him to do not necessarily what the strongest of his pre-existing motives would make him do unaided, but instead what he thinks it would make most sense for him to do, given all of his pre-existing motives and other relevant facts about his attributes and circumstances. The motivational power of the pre-existing motives that lose out in this process will be inhibited, whereas the winning motives will be strengthened and supported by the "managerial" motive of self-understanding (1996: 192, 196):

> [W]hat [the agent] does is going to depend on what he sees as making sense in light of [his circumstances, attitudes, and attributes]. His preexisting motives will be joined, and their balance potentially altered, by the very motive that leads him to think about them as clues to his next action, since that motive will incline him to do what those clues render it most intelligible for him to do. (2009: 132-133)

Strikingly, however, the structure of this explanation of the relevance of the notion of self-understanding for autonomous agency does not make any essential use of the notion of self-understanding itself. Velleman's explanation would work just as well if we said that the inclination with which agency is to be identified is an inclination towards cake: cake is the constitutive aim of agency. This inclination towards cake engages one's cake-conception—about how to get cake and what sort is best, etc.—altering the balance of one's other, pre-existing motives, inhibiting motives that one thinks would frustrate the aim of getting cake, and favoring motives that one thinks would help one to achieve that aim. Never mind that one's previously strongest motive was for kale—now one's previously weak motive for cake is strengthened and supported so much that one acts on that instead. Of course, nothing in these thoughts helps to render it plausible that cake really is the constitutive aim of agency, nor that positing a cake-inclination as an essential feature of agency would be especially helpful in understanding autonomy.

Velleman might protest that there is a crucial difference between cake-conceptions and self-conceptions, and the respective verdicts they help to yield about what to do. He writes:

> [I]n taking some move to be intelligible, the agent take[s] it to be intelligible in light of his circumstances, attitudes, attributes, *and his hereby taking it to be intelligible* […] [I]f he took something else to make more sense, the balance of his motives would be tipped in its direction – in

> light of which it really would make more sense, and as a result of which he would do it instead. (2009: 133, fn. 20)

Nothing parallel holds about cake. It is not true that if one took φ-ing instead of ψ-ing to be more conducive to cake, then the way in which the balance of one's motives would be altered would mean that φ-ing really would be more conducive to cake. One cannot simply make up one's mind about how best to get cake, whereas one *can*, Velleman seems now to be saying, simply make up one's mind about what it makes most sense for one to do, given one's circumstances, pre-existing attributes etc. This is what explains the fact that doing what makes most sense to one is a case of autonomous action, whereas pursuing cake as one sees fit is not (as such). One can literally "write" one's future by making up one's mind about what makes sense; whereas one can merely "read" from one's cake-conception what one is going to do in pursuit of cake, and then enact that conception. (For the reading/writing metaphor, see Velleman 2009: 132-133 *et passim*.)

However, this protest is problematic. If one's merely taking something to be most intelligible renders it most intelligible in fact (as the most recent quotation claims), and if one always acts on one's conception of what is most intelligible for one to do (as the quotation seems to also claim), then no action could fail by the standard of being most intelligible: any action one might perform will be the one that best achieves self-understanding. Since one could not, in this case, fail by the lights of the constitutive "aim," it could not be normative. Velleman must therefore retain the view that it is one's actual pre-existing attributes and characteristics that determine "what makes sense," at least to an approximation; one's own verdict of what makes sense can play at most a modifying role, and one's verdicts must be capable of being false. But then there is no relevant difference between the cake case and Velleman's view after all. One's views about how best to get cake *do* play a modifying role in how one is actually going to get cake, since they are what will prompt one, together with one's cake-inclination, to actually take steps towards getting cake.

Of course, this response depends upon one's cake-conception not being completely false. That conception can only lead one to cake (unaccidentally) if it includes, or can be updated to include, some

true beliefs about how to get cake. But the same is true of Velleman's view. One's initial self-conception about what would make sense cannot be completely false, for otherwise one's verdicts about what would make most sense in the light of that conception will simply fail to address any of the pre-existing motives whose motivational force that verdict is supposed to help alter. To take a toy example, suppose I have two motives: getting rich and being famous. But my self-conception is the completely false one that I want to live a humble and private life. My self-conception, together with my present situation, might lead me to think that what it would make most sense for me to do now is to spend all my dollars on a small farm near a quiet country town. How does my inclination towards self-understanding engage my actual pre-existing motives to alter their motive force? By inhibiting them, it seems. But given that the inclination towards self-understanding can play at most a modifying role, and yet its support is needed for autonomous action on any of one's pre-existing motives to ensue, I am simply stuck. I cannot act. And if one cannot act, then neither can one do anything to actually achieve self-understanding. A somewhat correct self-conception is a condition of achieving self-understanding, just as a somewhat correct cake-conception is a condition of achieving cake.

So it seems that my initial complaint against Velleman's rationale for his view stands. That rationale works as well for the cake view of agency as it does for Velleman's. Nonetheless, we might ask: is Velleman's view at least independently plausible?

We might be troubled by the extent of folk-psychological knowledge about one's own psychology that Velleman attributes to ordinary agents. Often what we do makes more sense in folk-psychological terms to others than it does to ourselves, and that is in part because knowledge of what one's own character traits are can be difficult to attain. We do not seem to have direct first-person access to facts about our character; and the common human tendency to see oneself in a favorable light imports distortions to one's self-observations that neutral observers' observations escape. However, it does seem that we usually have at least some insight into our desires, emotions, and so on. Perhaps that limited amount of self-knowledge is enough for the sort of partly true self-conception that Velleman's view requires for action to be possible.

A different worry is that it might seem pre-theoretically desirable to leave room for the possibility of occasionally acting completely "out of character"; but since Velleman's view had to confine the motivational potential of one's inclination towards self-understanding to a merely modifying role in order to restrict the range of actions it can "make most sense" for one to do, it is hard to see how Velleman could account for its possibility.

In fact, even milder cases of attempting self-reform look hard to account for on Velleman's view. One might of course *desire* self-reform—one might wish, say, to be generous rather than greedy—and this motive for self-reform would then be one of the motives whose influence one's inclination towards self-understanding can alter and modify. But as Velleman himself says, even a very strong desire need not be the one that wins out in this process. One's self-conception might not include that strong desire; and even if it does, one's verdict about what it makes most sense for one to do might not favor the fulfillment of that desire in particular. Given the "holistic Gestalt" character of the self-conception Velleman thinks determines our verdicts about what would make most sense for us to do, it is hard to see why a single desire for reform would be favored over the majority of one's attributes and characteristics that it wars against. Velleman claims that self-understanding must seek to resolve all conflicts; he regards it as a species of coherence (2009: 126, fn.12). But conflicts can always be resolved either way.

Velleman responds to something like this worry by saying that self-understanding favors resolutions that result in simpler, easier-to-understand selves rather than convoluted, hard-to-understand ones (2009: 32). But if so, then surely the quicker, simpler and easier way is to try to eliminate the desire for self-reform rather than to indulge it. Indulging a desire for self-reform would likely mean embarking on a long and arduous path of psychic conflict and duress, as one's old characteristics resist replacement.

Finally, it is not at all clear why self-understanding in folk-psychological terms must tend to resolve all conflicts. One might perfectly well understand what it is to be conflicted in the way one is: being conflicted is, folk-psychologically speaking, a very familiar way to be. So one might do *nothing* about it; indeed, given the intelligibility of such conflicts, one would be *stuck* doing nothing about it on

Velleman's view. In sum, then, the possibilities for self-reform, except by what looks like accident, look slim.

There is surely some plausibility to the simple thought that what we do tends to "make sense" by the lights of who we are, in folk-psychological terms. But this tendency need not be evidence of an underlying inclination towards what makes *most* sense by the lights of who the agent already is. The tendency of our actions may be simply to make *some* sense, in the only terms in which we can make each other's actions intelligible at all: in the folk-psychological terms that help to explain how agents' avowed reasons relate to their actions. In making sense of agents' actions at all, we thereby make sense of their actions as theirs. This thought places no constraints on what we can make sense of on a given occasion for a given agent. So it yields no reason to count actions that go against our established characteristics as thereby worse, or as done for worse reasons.

I have not, of course, said anything myself here to positively explain acting "out of character" or self-reform. But at least nothing in the view about autonomous agency that I have been defending in this chapter rules them out. Having some relatively stable practical dispositions can be a condition of the unity required for agency in general, even if it is possible to act completely out of character on occasion. To be sure, if acting "out of character" or in pursuit of self-reform inevitably introduces some measure of psychic conflict, then any account of the unity required for agency must be relatively modest if it is to accommodate their possibility. It is an interesting question exactly how this might be best achieved. At least part of the solution looks to be to deny not only that agency is essentially characterized by a disposition towards self-understanding, but also, as we already denied, that agency is essentially characterized by a disposition towards unity or coherence, just as such. (Cf. my remarks at the end of §3.3.4, above.)

This concludes my argument against Velleman's view. His rationale for the view is faulty; and the view itself has some counter-intuitive consequences when it is construed strongly enough to be capable of yielding practical norms. §3.5.1 argued that the sense in which spontaneous self-knowledge might be essential to agency likewise fails to yield practical norms. Hence our conclusion about autonomy

still stands. We have found nothing in the nature of autonomous agency to motivate the view that there is some metaphysical "must" about agency that can yield a normative "must." Autonomous agency can be exercised through practical dispositions that are contingent for the nature of agency as such.

3.6 CONCLUSION

I started this chapter by assuming that agency is essentially autonomous in some sense. I then defended a view of autonomous agency on which it is a matter of self-determination through contingent determinate forms of the determinable essence of rational agency. The practical dispositions through which each agent can exercise her autonomy are contingent for the nature of autonomous agency as such. This conclusion matters for the assessment of constitutivism, as it speaks against constitutivist attempts to locate some specific practical disposition in the very essence of autonomous agency, from which we might then derive practical norms. To support my conclusion about autonomy, I examined the central concepts figuring in constitutivists' own arguments in the area, in each case either accounting for the concept in non-constitutivist terms or rejecting its centrality to an account of autonomy.

If my conclusion about autonomy is right, then together with the arguments of chapter 2, it funds a comprehensively anti-constitutivist account of agency. For chapter 2 showed, in effect, that the fact that agency is essentially *efficacious* does not fund any constitutive norms either. Once we have accounted for how agency can be both efficacious and autonomous without importing constitutive norms, it is hard to see what other aspects of agency the constitutivist might try to exploit. In particular, the two chapters together look to amount to a *possibility proof* of agency without constitutive norms. If one's practical reasoning is both efficacious and autonomous, then that is enough for agency.

This does not mean, of course, that there aren't aspects of agency that I have not given anything approaching a detailed account of. For instance, I merely suggested, following Velleman, that spontaneous self-knowledge might be essential to autonomy; but I did not examine in detail either

whether we should really think this or what such knowledge might be. Instead I only showed in general terms that the idea of such knowledge does not, in any case, serve the constitutivist's ambitions. There is certainly much more to be said in that area, as there is in many others. But if my arguments are basically right, then my conclusions stand: neither autonomy nor efficacy helps to support constitutive norms, and so the nature of agency cannot help to yield an account of good practical reasoning and reasons.

4.0 IS THERE A METASEMANTIC ROUTE TO ETHICAL TRUTH?

Might metaethics allow us to discern the content of ethics? In particular, can first-order ethical disputes be resolved on the basis of arguments that operate purely at the metaethical level, begging no first-order ethical questions? This chapter argues that while we have reason to hope that a purely metaethical route to ethical truth exists, a promising class of strategies I will call *metasemantic* disappoints this hope. I also extend the argument to closely related strategies I call *metapragmatic*.

4.1 THE QUESTION AND THE STAKES: ETHICS, METASEMANTICS AND THE AUTONOMY OF ETHICS

Our question is whether purely metaethical, and in particular, metasemantic, arguments can vindicate a first-order ethical view. Let us start by defining the terms of our question, and by explaining why it matters what the answer is.

4.1.1 'Ethics'

I understand the category of the ethical quite broadly, as concerned with the general question *how one should live*, understood to be about the good lives of rational agents, just as such: agents whose characteristic activities, *qua* the sorts of agents they are, are acting for reasons and deliberating towards action. I assume that we, you and I, are rational agents in this sense. And so I assume that if there is a

correct answer to the question how one should live, this answer greatly concerns us. But I do not assume that only humans, say, are rational agents. And I make but one assumption about the content of the correct answer to the question how one should live: that if there are standards governing the good use of our capacity for rational agency—standards of practical reason—then those standards are central to the correct answer.

Central, but not necessarily exclusive of other aspects of the good life: there may be more to the good lives of actual rational agents than the good exercise of practical reason. Still, if there are no good reasons for action, and no such thing as good practical reasoning, then ethics is, perhaps surprisingly, mum about a central aspect of our practical lives:[160] there is ultimately no propriety or impropriety in the use of practical reason. This could leave ethics with something to say to actual rational agents, but nothing that concerns their proper conduct *qua* rational agents.

So in speaking of ethics, it will be standards of practical reason in particular that I am concerned with: what reasons, if any, is it good to act on (in a given context), and what, if anything, is it to reason well about what to do? Apart from this focus on standards of practical reason, I will understand the question how one should live in the broadest possible sense. It is a version of what Bernard Williams, in the first chapter of his (1985) *Ethics and the Limits of Philosophy*, calls 'Socrates' question'. The sense of Socrates' question how one should live is supposed to be neutral between various answers it might receive. In asking it, we are not to presuppose that it is best answered from the perspective of, say, commonsense morality, self-interest, or national interest (1985: 6, 19). Rather, Socrates' question precisely asks for a vindication of any system of putative demands that would constitute an answer to the question.

I wish to retain this generality: we are not to presuppose anything about the content of the region of ethics of interest to us. We are not, for instance, to suppose that the appropriate conception of practical rationality and practical reasons must speak especially to self-interest, instrumental rationality, morality,

[160] I do not here address skepticism about the very idea of practical reasoning, and of acting for reasons: I assume that there is rational agency. And I assume that being a rational agent is indeed a central aspect of our practical lives.

or economic considerations. What the standards of practical reason are, if there are any such standards, is precisely what is at issue in Socrates' question as I wish to understand it.

One clarification is in order at this stage. Williams' use of 'ethics' is narrower than mine. In my use, Socrates' question, understood to be about the standards of practical reason (if any), *just is* the central ethical question. In Williams' use, in contrast, the term 'ethical' denotes a type of concept or consideration, usually capable of serving as someone's reason for action, demarcated from the non-ethical, or as may be the case, counter-ethical, by its content; and the ethical is opposed from the outset to, say, egoistic, aesthetic, or economic considerations.[161] As a result, Socrates' question, neutral as it is supposed to be between possible answers to it, does not automatically receive an ethical answer in Williams' sense of 'ethics':

> If ethical reasons, for instance, emerge importantly in the answer, that will not be because they have simply been selected for by the question. (1985: 19)

For Williams, if the correct answer to Socrates' question tells us to act for ethical reasons, or to deliberate as the ethically good would—where these notions are understood independently of their figuring in the correct answer to Socrates' question—then this amounts to a vindication of ethics.[162] In contrast, I will take it that Socrates' question *just is* the central ethical question, and any correct answer to it to thereby counts as the central content of ethics.

I think both uses of 'ethics' are somewhat intuitive. We do often identify a concern or consideration as ethical because of its content. For instance, we identify something as the honest or dishonest thing to do; and we understand such thick ethical evaluations to be of a different sort, somehow, from (say) thick aesthetic evaluations of actions as graceful or clumsy. Williams' narrow use of 'ethics' is

---

[161] Except, of course, in the case of ethical egoists, who propound a theory of ethics (in Williams' sense of 'ethics') that is, from the point of view of commonsense ethical opinion, highly revisionary. See Williams 1985: 12; though as Williams remarks, it may not matter whether we call such a system an 'ethical' one.

[162] Cf. Williams 1985: 28-29. Williams himself is ultimately doubtful that such a philosophical vindication of ethics is available—but also, that it is as urgently needed as many philosophers seem to suppose.

perfectly legitimate.[163] But I will opt for the broader use here simply because the question how one should live seems to be the broadest question of obvious practical import to us. Indeed, Socrates' question is of obvious practical import to us precisely *because* of its breadth. In not being indexed to a potentially narrower standard, as is e.g. the question how one should live *in order to live aesthetically well*, Socrates' question escapes the further question why it should matter to us, in conducting our lives, what the answer to it is. If anything genuinely matters at all, it is surely the answer, if any, to Socrates' question. At any rate, if the question's obvious importance does *not* qualify it as ethical—and so any correct answer to it as the central content of ethics—then it is open to question just how concerned we *should* be about the content of ethics, in some narrower sense of 'ethics'.[164] (This last remark, of course, is of a piece with Williams' attitude, in the quotation above.)

In sum, I understand the ethical as centrally concerned with the question how one should live; and in asking this question, we are not to presuppose anything about the answer, except that it speaks (perhaps *inter alia*) to the good use of practical reason, if there is such a thing. The thought that the sense of Socrates' question is otherwise neutral between answers to it is of a piece with the hope of vindicating an ethical view without begging any substantive first-order ethical questions, in our broad sense of 'ethics'. Our question, then, is: if there are truths about what good practical reasoning and reasons are, is there a purely metaethical, and in particular, a metasemantic, argumentative route to these truths? §4.1.2 explains, in outline, what a metasemantic route to ethical truth would be. §4.1.3 explains what is at stake in the question whether such a route exists.

---

[163] Of course, there is an even narrower use of 'ethical', in which it is synonymous with the contents of what Williams calls 'the morality system' (1985: 174).

[164] This is exactly the sort of thought that commonly fuels worries about the rationality of morality. Philippa Foot 1972 treats of ethics, or "morality," as parallel to etiquette, in exactly the way that raises the question why one should care about its demands. And Foot is there skeptical that any categorical *ought* does legitimately back up the demands of *any* such system, whether it be a system of morality or of etiquette: the only imperatives there are are "hypothetical," in the sense that their demands have authority only for someone who already has a matching contingent allegiance to the system. In contrast, Foot's later work (1994, 2001) sees the ethical as much more closely tied to the standards of practical reason; Foot comes to regard it as a mistake to assume some antecedent conception of the standards of practical reason, to which the ethical is then supposed to match up, or else be damned.

4.1.2 What is a metasemantic route to ethical truth?

I will call a strategy *metasemantic* when it attempts to vindicate a target first-order ethical claim, for example, an instance of the schema "The fact that *p* is a reason for X to φ in circumstances *c*," by appeal to the conditions of understanding the meaning of a sentence that expresses that claim, or the meanings of its constituent terms. That is, metasemantic strategies appeal to the conditions of *semantic understanding* in an attempt to thereby vindicate a target ethical claim or view.[165]

How might the conditions of semantic understanding help to vindicate an ethical view? While there are different ways to attempt to spell out a metasemantic argument, here is the central idea of the one that, as I will argue, looks most promising: Among the conditions of understanding the target ethical claim is a commitment of some sort to that claim's truth; therefore one who understands the target claim cannot coherently doubt its truth. The metasemantic vindication of an ethical view stems from the view's indubitability to anyone who understands it—and since understanding a claim is a precondition of genuinely doubting it, this is just indubitability *tout court*.[166] Provided that indubitability is a guide to truth—a disputable premise, but as I will argue, a defensible one in the present context—the conditions of semantic understanding thereby yield a normatively non-question-begging argument for the truth of an ethical view.

In what follows, I will follow the literature in taking it that mapping the conditions of semantic understanding is equivalent to mapping the conditions of *possessing the concepts* that the constituent terms of the target claim express. (For a brief defense of the equivalence assumption in the context of discussing metasemantic arguments, see Williamson 2003: 272.) §§4.2-4.3 consider metasemantic arguments for ethical views in some detail, attempting to develop the most promising argument of this

---

[165] This is, at least prima facie, a slightly different use of 'metasemantic' from the use in which it signifies general concern with that in virtue of which expressions in a language have the semantic values they do.

[166] This leaves it open that one might *fail to accept*, and in this weak sense reject, a claim exactly *because* one fails to understand it. Supposing one fails to accept some truth of ethics in this latter sense, then is one nonetheless under its authority? And could one escape the authority of an ethical claim merely by somehow relinquishing one's understanding of its ingredient concepts? I come back to this briefly in §4.2.4, at p.158.

sort, but showing that such arguments fail. But what is at stake in deciding the fate of metasemantic arguments?

4.1.3 The stakes: the autonomy of ethics

What is at stake in the fate of metasemantic strategies is the nature of our intellectual reach to the credentials of any ethical view. Rejecting metasemantic strategies is a major step towards the conclusion that the only available sound arguments for ethical views are themselves ethically partisan, in the sense that they inescapably rely on first-order claims about the content of ethics that a skeptic could coherently doubt. While such arguments can be sound if their premises are true, their partisanship is disturbing. For such arguments could not be seen to be sound by anyone who does not already happen to share the right ethical view to at least some extent. If the only possible arguments in ethics were partisan in this way, we would be left with no intellectual response to skeptics about an ethical view, besides futilely repeating the arguments the skeptic rejects, or else engaging in some rousing rhetoric. Even if sound, arguments that are partisan in this way cannot shore up the confidence of anyone, including ourselves, who might be genuinely wondering about the credentials of a given ethical view.

Call the thesis that the only possible sound arguments in favor of any ethical view are partisan in the way described *the autonomy of ethics*. What is worrying about the autonomy of ethics is not just an idle worry about the possibility of skepticism: it is a worry about the nature of our own intellectual reach to the credentials of any ethical view. Metasemantic strategies are one way to attempt to transcend partisanship and to establish the credentials of an ethical view on an objective and normatively non-question-begging basis. For such strategies try to show that facts about the conditions of concept-possession can vindicate an ethical view without begging any normative questions against the skeptic in the process. Of course, arguments proceeding in purely first-order terms may also still be sound; but their soundness can now also be verified from *outside* of ethics, by appreciating the conditions of semantic understanding. However, if we reject metasemantic strategies, *and if there is no other sound way to resist*

the autonomy of ethics, then we must conclude that the only possible sound arguments for conclusions about how we should live are ethically partisan. Even if such sound arguments exist, their existence will be cold comfort to those genuinely wondering who, if anyone, does have the correct ethical view.

I take it that the autonomy of ethics is intuitively disturbing enough to give us some reason to hope that either a metasemantic strategy, or some other purely metaethical strategy of arguing for an ethical view, is sound. Unfortunately, if my arguments in this chapter are correct, then metasemantic strategies, as well as a closely related class of strategies I will call metapragmatic, disappoint this hope.

§4.4 returns to the consequences of rejecting metasemantic strategies in more detail. There we ask what other ways of denying the autonomy of ethics there might be, given the failure of metasemantic and metapragmatic strategies. To anticipate, I will argue that the only alternative strategy of resisting the autonomy of ethics is metaethical *constitutivism*, which attempts to ground ethics (in our broad sense) in the metaphysics of agency. I have argued against constitutivism in chapters 2-3: the nature of agency is ethically neutral. This chapter is, then, the final one in a dissertation arguing for the autonomy of ethics.

4.2 IN SEARCH OF A METASEMANTIC ROUTE TO ETHICAL TRUTH

Exactly how are metasemantic arguments for ethical views supposed to work, and why do they fail? Paul Boghossian's (2001, 2003a) metasemantic argument for objective *epistemic* reasons and norms of theoretical reasoning provides a good starting point for investigating how the conditions of concept-possession could justify conclusions about what one is entitled, or more strongly, obligated, to do or to believe.[167] On the basis of Boghossian's ideas, §4.2 first develops and rejects one argumentative schema that attempts to vindicate entitlements; and then develops a more promising schema that looks capable of vindicating both entitlements and obligations. §4.3 then considers whether this schema has sound

[167] Page references to Boghossian's articles will be to the reprints in his 2008b *Content and Justification*.

instances in the ethical case. I argue that it does not; the discussion further leads us to identify, and eventually to reject, two more related arguments I call metapragmatic. I also consider whether Wedgwood's (2007) general argument that the possession conditions of any concept must consist solely of rational dispositions, not irrational ones, can provide a metasemantic route to ethical truth.

My overall aim in §§4.2-4.3 is twofold: on the one hand, to seek the best argument that might provide the desired metasemantic vindication of an ethical view; and on the other, to show that such a vindication is not possible. So let us begin by asking: How can we get from conditions of concept-possession to entitlements or obligations?

4.2.1 Boghossian on the meaning-entitlement connection

We ordinarily assume that some types of theoretical reasoning are good, some bad, in the sense that some, but not all, inferential transitions lead us to justified beliefs, at least given that we are also justified in believing the premises.[168] That is, some inferential transitions are *entitling*: they entitle us to their conclusions. (Some may, further, be *obligating*; more on this below.) We also ordinarily assume that among the good rules of reasoning are, for instance, rules roughly corresponding to intuitively true claims about deductive logic such as

MPP *p*, *p*→*q* imply *q*.

But how may we justify an epistemic view that identifies certain rules of reasoning as good, others as bad? Boghossian worries that if we try to *argue* for an epistemic view, then we inevitably involve ourselves in some type of reasoning in the course of that argument. And for any such argument to entitle us to its conclusion—i.e., to the conclusion that a certain epistemic view is true—it seems that we must

[168] Some types of reasoning may be warrant-*generating*, not warrant-transferring. This is especially plausible in the case of practical reasoning. But nothing important turns on this distinction in the present context.

already be entitled to reason as we do.[169] If so, then no argument for the goodness of a type of reasoning will entitle us to its conclusion unless the rule we follow in mounting the argument is itself one that we are entitled to follow, and whose conclusions we are therefore entitled to trust. On the other hand, Boghossian finds no refuge from reliance on argument in the epistemology of reasoning. He considers, for instance, the proposal that we are *default* entitled in certain inferential transitions, but finds this proposal uninformative without an account of what default entitlement amounts to; and as soon as we start defending such an account and showing that it applies to some type of inference, we are involved in reasoning—reasoning whose deliverances, again, we are not entitled to trust unless we are entitled to that very reasoning (2001: 240-2).

It is plain, then, that if we must be entitled to a way of reasoning in order to be entitled to the conclusions arrived at by reasoning in that way, then there must be some type of entitlement to a way of reasoning that is not simply a further reason to think that that type of reasoning is good. For in order for that further reason to entitle us to the claim it putatively supports, the inferential transition from the reason to the claim must itself be a good one. Without a different type of entitlement to inferential transitions, a regress threatens. Indeed, severe skeptical problems threaten: inferential justification itself is in danger (2001: 246-7). This is where the conditions of concept-possession are supposed to help. Boghossian proposes that some inferential transitions are such that a disposition to engage in them is a condition of possessing some concept; and that the relevant inferential transitions are thereby entitling. His thesis is the following "meaning-entitlement connection" (2003a: 280):

---

[169] One might object to this requirement: for us to be entitled in the conclusion of an inference, isn't it enough that the inferential transition is in fact truth-preserving, and that we are entitled in believing the premises? But Boghossian thinks such an "externalist" condition on our entitlement to inferential transitions is insufficient, for there are truth-preserving inferential transitions to whose conclusions we are not entitled despite being entitled in believing the premises, because engaging in the inference in question would be "epistemically irresponsible." Boghossian's example is (validly but irresponsibly) inferring Fermat's last theorem from certain easily justified beliefs about whole numbers (2003a: 268-269). Against Boghossian, I doubt that what can be going on in such cases of epistemic irresponsibility is really *inference*. But I here put aside the plausibility of Boghossian's requirement, focusing instead on articulating his positive metasemantic idea that is designed to meet the requirement.

MEC Any inferential transitions built into the possession conditions of a concept are *eo ipso* entitling.

Why believe MEC? Boghossian writes:

> […] Suppose it's true that my taking A to be a warrant for believing B is constitutive of my being able to have B-thoughts (or A-thoughts, or both, it doesn't matter) in the first place [since constitutive of my possession of the concepts that figure in either A or B, or both]. Then doesn't it follow that I could not have been epistemically blameworthy in taking A to be a reason for believing B, even in the absence of any [independent] reason for taking A to be a reason for believing B? […] If inferring from A to B is required, if I am to be able to think the ingredient propositions, then it looks as though so inferring cannot be held against me, even if the inference is [one which I have no further reason to consider correct, or entitling]. (2003a: 279)

Boghossian's thought seems to be that one must be epistemically blameless in engaging in any inferential transition the disposition to which is constitutive of possessing some concept that is an ingredient of either one's premises or one's conclusion. Supposing that such blamelessness constitutes (at least defeasible) entitlement to engage in those inferential transitions, we need no further, independent reason for taking an inferential transition to be a good (i.e. entitling) one before we can trust it. We are entitled to engage in a concept-constituting inference simply because it *is* concept-constituting. If so, a certain epistemic view is true: namely, the view that the concept-constituting inferential transition in question, or perhaps better, the rule of reasoning we follow in making transitions of that type, is a good one.[170] This further corresponds to a truth about good epistemic reasons, via the following schema: A is a good reason for believing B if and only if the inferential transition from A to B is a good one (cf. Boghossian 2001: 235-6).

Boghossian is hopeful that at least MPP is concept-constituting and therefore entitling:

> [T]he idea that, in general, we come to grasp the logical constants by being disposed to engage in some inferences involving them and not in others, is an independently compelling idea. And the thought that, in particular, we grasp the conditional just in case we are disposed to infer according to MPP is an independently compelling thought. (2003a: 279)

If Boghossian is right, this would be a welcome result, since MPP-type reasoning, or some type of reasoning derivative from this type, is particularly likely to be involved in any argument for the goodness

[170] This may of course constitute only *part* of an epistemic view: there may be other concept-constituting, and in this sense basic, rules, as well as derivative rules.

of any type of reasoning (cf. 2001: 246). The epistemology of reasoning would be saved from crippling skepticism about the inferences needed to mount its own arguments.

To be sure, Boghossian's argument for MEC requires that the inferential transitions made in formulating the argument are themselves kosher. Boghossian admits this: where appeal to concept-possession is supposed to help is in vindicating such inferential transitions where they would otherwise remain suspect.[171] The rule employed in an argument is blamelessly employed (in that argument, as well as elsewhere), if it is concept-constituting (2001: 260-1). We might worry that MEC comes to the scene too late to play its desired role in undergirding inferential justification, if our justification for believing either MEC or the thesis that some rule or other is concept-constituting is itself inferential. However, if one's thesis that certain inferential transitions are concept-constituting and therefore entitling is true, then one *is* entitled to rely on those inferential transitions in supporting that very thesis (as well as elsewhere). The metasemantic proposal is therefore at least not obviously doomed.

Let us consider objections to Boghossian's argument, with the aim of teasing out exactly how metasemantic arguments must work in order to help us avoid the autonomy of ethics.

4.2.2 First objection: can we avoid the autonomy of epistemology/ethics?

As we saw, MEC can save inferential justification just in case it is *true*. This is so regardless of whether our argument for MEC itself is a good one; though our argument *may* also be good. Dialectical problems remain, however, for anyone wishing to justify a particular epistemic view. An epistemologist who genuinely doubts her entitlement to certain inferential transitions has no intellectual refuge from the worry that she might not, for all she can assure herself of, be entitled in safely concluding, with regard to any

[171] Boghossian sometimes puts this point in terms of rule-circularity: rule-circular arguments in accord with concept-constituting rules are good, where rule-circularity would otherwise be suspect (2001: 245-6, 260-1). My formulation concerns reliance on rules of reasoning more generally, including cases where two or more different rules might be mutually supporting. (This is not strict rule-circularity.) Since Boghossian's main concern is with vindicating our reliance on types of reasoning in general, this does not matter for the main point. Cf. 2001: 247.

inferential rule, that it is indeed concept-constituting, nor that being concept-constituting entails being entitling. She can only hope that her reasoning in support of those conclusions is itself entitling, and that her conclusions are true. As epistemologists, then, we cannot hope to vindicate an epistemic view without *placing faith* on some aspect of just such a view. Of course, if our thesis that MPP is concept-constituting and therefore entitling is true, then a would-be skeptic about MPP cannot in fact doubt MPP without being implicitly bound to be disposed to reason in accord with it: for in so much as having the concept *conditional* required to formulate one's doubts, one has that disposition. Yet the epistemological skeptic would, of course, continue to deny either MEC or the claim that MPP is concept-constituting, thereby denying that she is in the predicament we claim she is. (Doubting MPP's goodness, she might consider it a counterexample to MEC.)[172]

The curious result is that, although a sound metasemantic argument for MPP *would* ensure that even the skeptic is implicitly committed to reasoning in accord with MPP, and so (if MEC is correct) implicitly committed to the view that that MPP is entitling, we cannot tell, without placing faith in our own reasoning, whether the skeptic is in this predicament or not. We cannot tell, without placing faith in our own reasoning, whether we have "begged the question" against the skeptic, in the sense that we have claimed something she can coherently reject.[173] And so we cannot tell whether we have avoided the *autonomy of epistemology*: the thesis that there is no sound argument for an epistemic view that begs no

---

[172] It is a further question why the necessity of being disposed to reason in accord with MPP in order to formulate doubts about MPP's goodness would imply that the skeptic cannot coherently doubt MPP's goodness. It certainly *seems* coherent to raise doubts about the goodness of a way of reasoning even as one is in fact disposed to reason in that way (cf. Williams 1985: 28). For example, most humans are disposed towards some fallacious types of reasoning, and it makes sense to doubt the goodness of those types of reasoning. The metasemanticist might respond that the reason why the doubts make no sense in the case of MPP is precisely that one cannot even have the concepts required to formulate one's doubts without being disposed to reason in accord with MPP, and thereby implicitly committing oneself, via MEC, to MPP's being entitling; whereas the same is not true in the case of fallacious types of reasoning. But the skeptic who denies MPP's goodness in part because she denies MEC should not, by her own lights, be moved by this. Given that Boghossian's question is, in effect, precisely *which lights are good*, the dialectical problems remain.

[173] There is of course a more stringent notion of "begging the question," on which even the dialectical impasse described counts as begging the question against the skeptic. Thanks to Kieran Setiya for pressing me to clarify these issues.

first-order epistemic questions against skeptics.[174] (Of course, since "the skeptic" might simply be ourselves wondering about the credentials of our own epistemic views, what we cannot tell is whether we in fact have a good epistemic view or not.)

This may make us worry that, likewise, we will not be able to judge whether metasemantic arguments can help us avoid the autonomy of ethics. But in fact no parallel complications arise in the ethical case. The dialectical difficulties arose in the epistemic case because, in so much as mounting the metasemantic argument, we are forced, in our reasoning, to rely on some aspects of the epistemic view we are trying to vindicate. But any ethical view we might attempt to vindicate concerns the goodness of rules of *practical* reasoning, rules for reasoning towards *action*; whereas in mounting the metasemantic argument itself, we are engaged in theoretical reasoning, reasoning towards a *belief* about the goodness of an ethical view. So we cannot be suspected of begging the question against ethical skeptics in merely relying on certain rules of theoretical reasoning in mounting the metasemantic argument. If there is going to be question-begging involved, it will have to be of a rather straightforward sort: one of the premises will have to be smuggling in assumptions about the content of the right ethical view, assumptions that a skeptic could coherently reject. Hence, if metasemantic arguments for ethical views can avoid premise-circularity, and they are otherwise good, then they can also straightforwardly help us avoid the autonomy of ethics, in a way that even the ethical skeptic could judge to work.

---

[174] As a result of similar reflections, Boghossian concludes, more strongly, that even sound metasemantic arguments cannot "refute" skeptics: and that our inability to refute skeptics therefore has no epistemological significance (2001: 265). However, if our metasemantic argument in support of MPP is sound, and the skeptic has to have the concept *conditional* in order to formulate her doubts about MPP, then the skeptic is implicitly committed to the claim that MPP is entitling, the very claim she attempts to doubt. So doubting MPP is incoherent. This is of course refutation only by our lights. But if our lights are in fact good—something the metasemantic argument at least makes possible—then isn't the refutation also good?

4.2.3 Second objection: counterexamples to MEC

Pejorative concepts such as *boche*, and made-up logical connectives such as *tonk*, have been proposed as clear-cut counterexamples to the meaning-entitlement connection (MEC) Boghossian posits. Arthur Prior (1960) stipulated that, to possess the connective *tonk*, one must accept, in the sense of being "willing to infer" in accord with, the following introduction and elimination rules:

$$\frac{A}{A \text{ tonk } B} \qquad \frac{A \text{ tonk } B}{B}$$

But as Prior observes, it would clearly be absurd to conclude that tonk-introduction and tonk-elimination thereby entitle us to their conclusions. In a similar vein, it may seem plausible (at any rate, it does to Boghossian; 2003a: 280) that, to possess the concept *boche*, one must accept the following introduction and elimination rules:

$$\frac{x \text{ is German}}{x \text{ is boche}} \qquad \frac{x \text{ is boche}}{x \text{ is cruel}}$$

Yet it seems wrong to conclude that one is thereby even defeasibly entitled in inferring that x is cruel from a (justified) belief that x is German.

Boghossian's own response to such counterexamples is to restrict MEC to only some concepts.[175] (He also considers denying that 'tonk' expresses any non-defective concept.)[176] It strikes me, however, that such stipulated conditions of concept-possession are blatantly implausible, and thus do not constitute genuine counterexamples to MEC. If there is a concept *tonk* that Prior's argument is about, then in order to even grasp the argument, it seems that we must grasp, in some way, the concept he aims to introduce. Yet it is absurd to think that anyone who can follow Prior's argument is thereby willing to infer

[175] In particular, Boghossian holds that where a conditionalized version of a concept is available, the unconditionalized version is not entitling (2003a: 281-5).
[176] 2003a: 281.

absolutely anything from absolutely anything, as they would be, were they willing to infer in accord with tonk introduction and tonk elimination.

Likewise with *boche*. Even if xenophobes use 'boche' in something like the stipulated way, understanding this fact about how xenophobes use 'boche' is surely itself enough to understand the concept *boche*; and in order to understand such facts about how xenophobes use 'boche', we need not *ourselves* use 'boche' in that way, or at all. Non-xenophobes can express their semantic understanding of 'boche' by *mentioning* the word in the course of spelling out the objectionable conversational implicatures involved in xenophobes' uses of it (cf. Williamson 2003: 267-8). In any case, it is surely exactly because one possesses semantic understanding of 'boche' that a non-xenophobic person would studiously *avoid* using, as opposed to mentioning, 'boche'.[177]

So the traditional counterexamples strike me as poor challenges to MEC. The inferential transitions that such putative counterexamples identify as patently unentitling are not even concept-constituting. MEC is still alive; and if some ethically relevant concept does turn out to have, among its possession-conditions, some type of commitment to specific patterns of practical reasoning, then MEC might vindicate those patterns as good (in the sense of entitling).

4.2.4 Third objection: entitlement to concepts and begging the question

There are, however, different problems with MEC. Even if some type of reasoning R is a condition of possessing some concept C, why should this yield anything but a *conditional* entitlement to engage in R-type reasoning, unless one is, in addition, entitled to possess C? It is implausible as a general principle that a way of reasoning is entitling just because it is a condition of X: it matters what X is. For example, suppose X is the belief "B follows from A." It seems to be a condition of one's genuinely believing this that one be at least in some way willing to infer B from A. But this is plainly insufficient to entitle one to

[177] As will transpire in §4.3, however, there are *some* uses available even to non-xenophobes.

that inference. We clearly need to restrict X in *some* way, if being a condition of X is going to render an inferential pattern entitling.

Concept-possession seems to be a promisingly restricted "X," since not just any inferential pattern is plausibly a condition of concept-possession. But merely achieving *a* restriction is not enough on its own to explain why being a condition of concept-possession should render a way of reasoning anything but conditionally entitling. For consider that not just any way of reasoning is, we hope, a condition of being a fallacious or biased reasoner; and yet being a fallacious or biased reasoner is not plausibly the sort of thing that can yield entitlements to conditions of it.

So we need to say what is special about concept-possession besides the fact that not just any inferential transition is plausibly a condition of possessing some concept. One proposal, already mentioned, is that one might be entitled to concepts. Attempting this line, we might amend MEC to include a hidden condition to the effect that one is always at least defeasibly entitled in possessing a concept, whatever the concept, and attempt to spell out MEC more fully as follows:

The Entitlement Argument

(1) To possess the concept C, one must be disposed to reason in accord with R.
(2) One is always at least defeasibly entitled in possessing a concept, whatever the concept.
(3) If one is at least defeasibly entitled in possessing a concept, then one is (thereby) at least defeasibly entitled in satisfying the conditions of possessing that concept.
(4) So, one is always at least defeasibly entitled in satisfying the conditions of possessing C.
(5) So, one is always at least defeasibly entitled in being disposed to reason in accord with R.

(5) states an unconditional, if defeasible, entitlement to a disposition to reason in accord with R. How do we get from (5) to the truth of the epistemic view that R is an epistemically good rule, in the sense that it entitles one to its conclusions (at least if one is entitled in believing the premises)? We must assume that being defeasibly entitled in being *disposed* to reason in accord with R also entitles one, at least defeasibly, to any *instance of actually following* R.[178] Given this assumption, we get the result that R is a rule that

[178] Williamson doubts this assumption (2003: 252-6). Discussing these doubts will take us too far afield here; but in brief, they can be resisted by insisting that genuinely good dispositions are discriminating in such a way that they

one is always at least defeasibly entitled in actually following.[179] If being entitled in following R secures the claim that R-type reasoning entitles one to its conclusions (given one's entitlement to one's premises), then R is a good rule in the specified sense. Hence the epistemic view that R is a good rule is true.

This argument looks quite promising. One striking feature of it is that one need not actually *have* the concept C to be entitled to reason in accord with R. It is enough that one is entitled to have the concept, and that this, together with the conditions of having it, implies one's entitlement to R. The conclusion that R is a good rule in the sense of being entitling therefore has universal application. (Or at least, it is universal over everyone so much as capable of possessing C.)

Another striking feature of the argument is that it does not depend on the assumption that C is an ingredient in R, or in the thoughts one thinks in the course of any actual episode of reasoning in accord with R. (For example, when R is MPP, that one inevitably thinks thoughts involving the concept *conditional*.) C could be any concept. Nonetheless, it is hard to see how a premise of type (1) could be defended unless C and R are internally related somehow. And apart from C's being somehow an "ingredient" in R, or in the thoughts one inevitably thinks in the course of any actual episode of following R, there is only one type of "internal" connection between R and C that looks remotely plausible. This is the case in which a disposition to follow R is a condition of possessing C simply because that disposition is a condition of being a thinker at all. The claim that a disposition to follow R is a condition of being a thinker would, however, serve as a premise in a more direct *constitutivist* vindication of R: roughly, since following R is part of being a reasoner, just as such, then following R well is part of what it is to reason well. (I have dealt with constitutivist arguments for ethical views elsewhere.[180]) The further link to R's being a condition of possessing C would then be a detour. Since my aim here is to engage metasemantic

---

equip their possessor to correctly deal with any instance to which they truly apply. Given this condition on the relevant kinds of disposition, the argument goes through.

[179] It would not do to change the argument, so as to get to this result, by changing premise (1) to the claim that any actual instance of following R is a condition of possessing C. That would be far too implausible a claim.

[180] Chapters 1-3.

arguments only as they might vindicate ethical views even where constitutivism fails, I ignore this type of case.[181]

With these clarifications in place, what to make of the Entitlement Argument? (3) looks plausible. (1) is obviously a key premise, but I leave its evaluation, in the cases relevant to ethics, until §4.3. What about (2)? Does the notion of entitlement to concepts even make sense?[182] Can a parallel argument be run in the ethical case? And does assuming entitlement to concepts make the argument question-begging in exactly the way that we are trying to avoid in avoiding the autonomy of ethics?

We might try to understand the idea of *epistemic* entitlement to concepts through the idea of its being epistemically blameless to possess a concept. Possessing a concept does not seem to be epistemically blameworthy or objectionable in and of itself, unless the concept is not a genuine concept, or is somehow defective. If anything, expanding one's conceptual repertoire expands the range of facts one can know, and so enhances one's epistemic capabilities.[183] This does not require us to view *employing* just any concept in its usual way as an epistemic advance: pejorative concepts, whose ordinary use signals the speaker's prejudice, are plausibly a case in point. Nonetheless, *possessing* the concept enables one to e.g. refer to it, and to its usual use by groups of people, thus expanding the range of facts one can know. So possessing a concept seems to be epistemically blameless, and in fact a type of epistemic advance, in and of itself. If being epistemically blameless in doing X amounts to being at least defeasibly entitled in doing X, then it follows that one is always at least defeasibly entitled in possessing a concept, whatever the concept. (Boghossian himself tends to equate blamelessness and defeasible entitlement; 2003a: 279-80.)

---

[181] In fact I have only discussed arguments for ethical views based on conditions of *agency*—of being a *practical* thinker. I have not discussed possible arguments for ethical views based on the conditions of being a *theoretical* thinker. But such possible arguments do not seem very promising. What in being a theoretical thinker could require one to have a particular practical disposition, except (i) the general condition that to be a theoretical thinker one must also be a practical thinker, an agent; together with (ii) a view of agency on which agency requires the specific practical disposition in question? While (i) is a possibility, I have repudiated (ii) in repudiating arguments for ethical views based on the conditions of agency.

[182] I thank Kim Frost and Kieran Setiya for pressing me to think hard about what entitlement to concepts might be. The following line of argument stems from my attempts to answer their concerns about this notion.

[183] A thought along these lines may be part of what Ralph Wedgwood has in mind in remarking that possessing a concept is always an *ability*, not a liability (2007: 168-9). I discuss Wedgwood in §4.3.3 below.

Supposing there is something to this, we might try a parallel argument in the ethical case. To do so, we must employ a notion of *practical* entitlement or permission, since a rule can be practically good, in the sense that it genuinely entitles us to the actions it prescribes, only if it is practically entitling. (A good rule may of course also be obligating; more on this below.) So the entitlement at issue in (5), the entitlement to be disposed to follow a practical rule, must be, somehow, of a practical, or at least of a patently practically relevant, sort. It is hard to even understand what it could mean to deny this—to say that *epistemic* entitlement to follow a practical rule R renders R *practically* entitling—*unless* this was just a strange way of expressing the view that we have some good reason to believe (and so epistemic entitlement to believe) that R is a good practical rule (in the sense of being practically entitling). But it is precisely the metasemantic argument that is supposed to provide us with such a reason to believe, of some R, that it is a good practical rule. And the Entitlement Argument would equivocate on 'entitlement' if it read it in (5) as practical entitlement, but in (2), (3) and (4) as epistemic entitlement. To make the argument valid, then, we instead have to read premises (2), (3) and (4) all as employing the same notion of practical entitlement.

But what is practical entitlement to a concept? We might suggest that, just as merely having a concept did not seem epistemically blameworthy, and indeed seemed like an epistemic asset, likewise merely having a concept does not seem practically blameworthy, and could be a practical asset. In having a concept, one is capable of e.g. appreciating potential grounds for action that those who lack the concept are not capable of appreciating. If practical blamelessness amounts to at least defeasible practical entitlement, then perhaps we can be practically entitled in possessing a concept.

The trouble is that it is very hard to see what practical entitlement to possess a concept might be. It seems that it must be some sort of entitlement by the lights of the standards of practical reason. But those standards are about good practical reasoning and reasons: they do not seem to even apply to having a concept. What might it be to have good or bad practical reasons for having a concept? And if I had good reasons *against* having a concept and no reasons for having it, ought I to relinquish my grasp of it? If I ought to, I wouldn't know how to.

One suggestion is that the standards of practical reason might approve of a concept in rather like the way in which they can approve of a disposition of reasoning: a disposition of reasoning is not something it is easy to acquire or relinquish, certainly not at will; but nonetheless there are good dispositions of practical reasoning, namely, those that are dispositions to engage in types of inferential transition that are good. Encouraged by this, we might suggest that a concept is a good one to have, by the lights of the standards of practical reason, if having it involves one's having a disposition to engage in good kinds of practical reasoning. Indeed, something like this seems to be the only way to make sense of practical entitlement to concept-possession: we must suppose that possessing a concept C consists, at least in part, in engaging in certain *practical uses* of it, or in being disposed towards those uses; only then could the standards of practical reason apply. But that means that we must presuppose, in defending the idea of entitlement to C, that an instance of premise (1) is true of C. And that renders the suggestion that we are entitled to possess C question-begging in exactly the way we are attempting to avoid. In supposing we are entitled to possess C, we would be presupposing that we are entitled to R-type practical reasoning.

It does not help to object that we might be entitled to *all* concepts, and that our entitlement to C, and so to its associated types of reasoning, merely follows from that. If entitlement to concepts does not make sense without supposing that possessing a concept consists in part in a disposition to engage in a certain type of practical reasoning, then in supposing that we are entitled to all concepts we can only be supposing that we are entitled to certain types of practical reasoning, namely, those constitutive of possessing concepts. If there are types of practical reasoning that are not concept-constituting, and that do not merely derive from the kinds that are, then the supposition that we are entitled to the concept-constituting types begs the question against the others, ruling them out without explanation. If, on the other hand, there are no types of practical reasoning that are not either concept-constituting or derivable

from concept-constituting types, then MEC rules nothing out, and so cannot in any case help us sort out the good from the bad types of reasoning.[184]

Finally, if, in supposing that we might be entitled to concepts, we are forced to presuppose that we are entitled to concept-constituting ways of reasoning, then that does not help us spell out or defend MEC: we have merely restated MEC in one of our premises. Hence we are left with no answer to the question with which §4.2.4 began: namely, why having a concept entitles us to the ways of reasoning that are supposed to be its possession-conditions, while e.g. having a bias or prejudice do not.

In sum, then, the Entitlement Argument is caught in a dilemma. Either it cannot make sense of the notion of entitlement to concepts, and so cannot yield anything but conditional entitlements; or it can make sense of entitlement to concepts, but in doing so it begs the question in favor of the types of reasoning, if any, that true instances of its premise (1) would involve. No matter what the concept C and its possession-conditions, then, the Entitlement Argument cannot yield a sound defense of an ethical view in a way that avoids the autonomy of ethics.[185]

It is worth noting that even if we conclude that the notion of entitlement to concepts makes no sense—perhaps because no concepts have the types of possession-condition that could make sense of that notion—we can nonetheless accommodate the intuition that having concepts is both epistemically and practically blameless. For blamelessness need not amount to entitlement. I might be blameless in having a concept, but I might also be blameless in having a pimple: X's blamelessness by the lights of a set of

---

[184] Could there be only one type of practical reasoning, and could it be that that type's being concept-constituting at least explains its being entitling? I argued in chapters 2-3 that there are several types of practical reasoning, not all of which can simultaneously be good. In any case, as I will presently explain, it is still doubtful what sort of an explanation of entitlingness MEC can provide.

[185] Might Boghossian's suggestion that only conditionalized concepts are entitling help? Not in the case of practical rules, it seems. Conditionalization is supposed to help, for Boghossian, by not prejudging whether anything exists that falls under the concept. For instance, a conditionalized version of *neutrino* says that *if* certain conditions in fact obtain as regards something, x, then x is a neutrino. This leaves it appropriately open to empirical investigation whether neutrinos exist. (The concept conditional, however, cannot be conditionalized, since it is presupposed in any instance of conditionalization; Boghossian therefore takes it that the possession-conditions of *conditional* can be entitling even where the possession-conditions of other unconditionalized concepts, concepts that do have conditionalized versions, are not (2003a: 282-285).) But our worry in wondering about the goodness of practical rules is precisely a worry about what the conditions of being a good reason, or a good disposition of practical reasoning, are. A conditionalized version of e.g. the concept good reason might be the outcome of our inquiry, but it cannot aid in an argument towards that outcome.

standards S need not come in the guise of X's being something we are entitled to by the lights of S, but instead, in the guise of X's not being condemned by S simply because nothing in S even intelligibly applies to things of X's kind. Indeed, we can even admit that having concepts is a kind of epistemic and practical asset in the ways outlined. So is having eyes and ears and microscopes. It still does not follow that we must be practically or epistemically entitled to these things.

Perhaps a metasemanticist should respond to all of this by conceding that the notion of entitlement *to* concepts is a category mistake, or at any rate not what MEC needs: we went wrong in so much as attempting to explain MEC by appeal to entitlement to concepts. Perhaps MEC is better construed on analogy to perception. Just as the normal operations of our perceptual faculties entitle us to their outputs—at least we hope they do—likewise the normal operations of concepts in cognition, namely, their involvement in reasoning of some type R, entitle us to the outputs of such reasoning. Concepts yield unconditional entitlements to their possession conditions not because we have some further entitlement to those concepts, but rather, because having a concept is entitl*ing*. However, while this may be a felicitous way of describing the idea behind MEC, it again leaves us with no answer to the question with which we started: why having a concept renders the ways of reasoning that are supposed to be the concept's possession-conditions entitling, while e.g. having a bias or prejudice or a megalomaniac tendency does not seem to render the ways of reasoning that are their possession-conditions entitling. MEC still lacks a good defense.

We can, however, formulate an alternative metasemantic argument schema that looks more promising than the Entitlement Argument; and it can take us to both entitlements and obligations.

4.2.5 An alternative metasemantic argument: the skeptic's commitments

We often expect practical norms, at least, to be obligating, not only entitling.[186] Sometimes we *ought* to do something. How might a metasemantic argument take us all the way to obligations? To be sure, there is a conceptual connection between the notions of entitlement and obligation. What is not permitted is forbidden, and one is obliged to avoid the forbidden. And it is, further, plausible that if there is a choice between inferential patterns to which one is entitled and ones to which one is not entitled, then one is obliged to follow some pattern to which one is entitled—if indeed one must reason at all. But we cannot exploit these connections in an attempt to strengthen entitlements to obligations without a good defense of entitlements to begin with. We need a different argument.

The idea that one is always *obliged* to possess a concept, whatever the concept, seems even less plausible than the idea of entitlement to concepts, and seems likely to encounter exactly the same problems with question-begging. There is, however, a more direct route from conditions of concept-possession to obligations, by appeal to the incoherence of skepticism about certain sorts of concept-constituting rule.

The Obligation Argument

(11) Anyone possessing the concept C must in some way take RO to be genuinely obligating, at least implicitly.
(12) Any would-be skeptic about RO must possess C: one could not even formulate one's doubts about RO without possessing C.
(13) So, any would-be skeptic about RO must take RO to be genuinely obligating, at least implicitly. (From 11 and 12)
(14) So, it is incoherent to doubt RO's being genuinely obligating: skepticism about RO is self-defeating. (A restatement of 13)
(15) If it is incoherent to doubt that *p*, then *p*.
(16) So, RO is in fact a genuinely obligating rule.

[186] Even Kant's categorical imperative yields obligations, via the thought that one is obliged to avoid what one is not permitted to do, and obliged to do what one is not permitted to omit. More about the entitlement-obligation connection below.

Since my interest is with the possibility of vindicating ethical views, let us discuss this argument solely as it concerns putative *practical* rules RO, and so ethical views to the effect that RO is genuinely obligating. Of course, the same argument schema promises to vindicate entitlements as well as obligations, if we just substitute 'entitling' for every instance of 'obligating'. But I will frame the discussion in terms of obligations. Several observations and clarifications are in order.

First, by saying that RO is "genuinely obligating" I mean not that RO *purports* to be obligating—not that RO *says* something to the effect that one ought, perhaps *ceteris paribus*, to do φ in circumstances *c*—but rather, that RO is in fact genuinely obligating. Likewise, when premise (11) and, subsequently, sub-conclusion (13), employ the notion of someone's "taking" RO to be genuinely obligating, more is intended than that someone takes it that RO *claims* that one is obligated (perhaps *ceteris paribus*) to φ in *c*. (Indeed, if RO is imperatively formulated, in roughly the form "In *c*, do φ!", it will state no claims.) Instead, when someone takes RO to be genuinely obligating, she takes it that she ought to do what RO tells her to. She takes RO to be authoritative.

But second, what exactly is it to "take" RO to be authoritative? One straightforward way of "taking" RO to be authoritative or genuinely obligating is to believe that it is. So there is at least one available construal of the relevant "taking" that makes sense. Of course, there are also problems with this construal. We might think it too intellectually demanding to be plausibly a condition of concept-possession. Boghossian himself works with the idea of being *disposed* to reason in accord with a rule precisely because of this. He thinks that while "accepting" MPP is in some sense a condition of having the concept *conditional*, interpreting such "acceptance" in terms of a belief in, say, the rule's validity would be implausible, since it seems that one might have and reason with *conditional* without having the concept *validity* (2001: 243). The trouble in the present context, however, with working with the notion of a disposition to reason in accord with RO, is that a skeptic about RO's authority might admit that her ability to formulate doubts about RO entails that she has some concept C, which in turn entails that she is disposed to follow RO. But why should she also admit that having this disposition entails that she

implicitly "takes" RO to be genuinely obligating? It certainly seems that we can coherently question a rule's authority, or our own right to follow it, even as we continue to be disposed to follow it.[187]

These problems with the dispositional construal of "taking" may be soluble, but they are very difficult, and beyond our scope here.[188] On the "implicit belief" construal, we at least have a way of explaining both what it is to "take" RO to be genuinely obligating, and why the skeptic's doubts about RO's authority are incoherent. If the "taking" in question is just a belief, then the would-be skeptic about RO attempts to contradict a belief that she inevitably has, a belief that RO is genuinely obligating; and since believing is believing to be true, the skeptic cannot also coherently hold that RO is not genuinely obligating. Even in so much as pretending to suspend judgment about RO's authority, she deceives herself.[189]

Finally, third, premise (12) tells us something about the concept C: C must be either (i) an "ingredient" of RO or of any possible application of RO in some way, so that one could not even entertain thoughts, including doubts, about RO without possessing C; or else (ii) C must be the concept *authority*, or perhaps *obligation*, so that when one attempts to doubt that RO is authoritative or obligating, one's possession of these concepts itself somehow commits one (as per premise 1) to taking RO's claims to be authoritative. There is no relevant third option on which (iii) the concept C is some concept completely unrelated to either RO or to concepts such as *authority* in terms of which one might formulate doubts about RO's ethical import. For again, it seems that the only way in which RO could be a condition of possessing C in cases of type (iii) is if both a commitment to RO's authority and possessing C were independently necessary aspects of being a thinking being, so that the one always entails the other, albeit

---

[187] On this point, see Williams 1985: 28.

[188] Sebastian Rödl (2010, 2007: chs.2-3) tries to explain why one must, in simply reasoning a certain way, take it that one's reasons are good ones, without this being a matter of one's having a further belief about the normative status of one's reasons. Nor does Rödl's account make use of the notion of a disposition. But the account is very difficult and I don't fully understand it. The issues here concern whether and how acting for reasons must take place "under the guise of the good." For a denial that it must, see Setiya 2007, 2010.

[189] This concession to straightforwardness will not matter for my argument below. Our present purpose is just to see how something like the Obligation Argument might be made to work, so as to see why its key premises are worth engaging.

mediately. And this in turn would suggest the more direct strategy of argument that, since any skeptic is a thinking being, and any thinking being is committed to taking RO to be genuinely obligating, it follows that any skeptic is so committed; therefore skepticism about RO's obligatingness is incoherent. Any focus on possessing C would be a detour in such an argument; and again, such an argument would be an instance of a constitutivist strategy—a strategy with which I have dealt elsewhere.[190]

With these clarifications in place, what to make of the Obligation Argument? Its crucial premises are clearly (11) and (12). Whether there are any true ethically relevant instances of these will be discussed in §3. But premise (15) is also controversial. It is of course trivially true that any claim whose negation is incoherent must be true; but why think that the necessary commitments of skeptics, or perhaps of thinkers more generally, are a good guide to what is in fact true? It might be incoherent to doubt that *p*, but this might be only because of the nature of doubt, or of doubters, not because of *p*'s putative truth. Indeed, if the incoherence of doubting *p* had something to do with *p*'s truth, then this would have to be because *p* is an analytic truth. And if *p* is an analytic truth, then surely that fact is already vindication enough for the view that *p*: why does it matter, in addition, that the skeptic cannot but believe it?

In response, we might suggest that metasemantic arguments *are* in fact ways of showing that certain (perhaps unobvious) analytic truths hold. But Boghossian himself rejects the notion of "metaphysical" analyticity, whereby a sentence 'p' is metaphysically analytic if and only if it is true wholly in virtue of its meaning. What he defends is "epistemic" analyticity, the idea that grasp of meaning yields entitlements.[191] What metasemantic arguments are supposed to spell out is exactly *how* grasp of meaning yields entitlements—or, as in the argument I am proposing, obligations. The conclusion of any metasemantic argument, to the effect that a certain rule R is a good one (whether by being entitling or by being obligating), is not supposed to be true in virtue of its meaning; rather, it is supposed to be true because of facts, recorded in the premises, about the conditions of understanding meaning. Exactly how those facts make those conclusions true is what metasemantic arguments attempt to spell out.

---

[190] Chapters 1-3.
[191] See Boghossian 2003b for this distinction.

If this is right, however, then we have not yet responded to the original worry about (15). Why think that indubitability is a guide to truth? Here is a defense of (15), in the context of the Obligation Argument. If it is incoherent to doubt that RO is obligating, then we cannot coherently conclude that it is not. Indeed, if premise (11) is true, and if having C is required for even suspension of judgment about RO's obligatingness, then we could not even coherently suspend judgment. We can only conclude that *p*. But now, if it were nonetheless the case that not-*p*, then there would be an ethical fact that no-one could coherently accept. The hypothesis that some ethical facts are like this amounts to the hypothesis that some ethical facts are in principle unknowable. And that is a skeptical hypothesis far more severe than the simple doubt that some practical rule RO might not be authoritative: this new hypothesis embodies a quite sweeping epistemological skepticism about ethics, whereas the skeptic of the Obligation Argument means only to press a rather ordinary, if still difficult to answer, doubt about the authority of a given practical rule. So while (15) may be difficult to defend as a fully general principle, it may be fine in the context of an ethical Obligation Argument, given the tacit proviso that the Obligation Argument, like most arguments, relies on the falsity of a sweeping epistemological skepticism.[192]

This is not conclusive, of course. But I leave it to would-be proponents of metasemantic arguments for ethical views to further respond to concerns about (15), or else to propose a better metasemantic argument, and one that takes us from claims about the conditions of concept-possession not only to entitlements, but also to obligations.

4.2.6 Conclusion

In §4.2, I first argued that, while even a sound metasemantic argument for an epistemic view is insufficient to assure us that we have avoided the autonomy of epistemology, the prospects are better that

[192] Of course, in the epistemological context, such a defense of (15) might be less plausible, at least given Boghossian's worry that without a metasemantic vindication of an epistemic view, we cannot see how objective epistemic reasons (and consequently knowledge of them) are even possible.

a sound metasemantic argument for an ethical view could straightforwardly assure us of the falsity of the autonomy of ethics. I then investigated in detail how metasemantic arguments work. Interrogating Boghossian's proposed meaning-entitlement connection, MEC, I developed but ultimately rejected the Entitlement Argument as a way of spelling out the metasemantic idea. I then proposed a different metasemantic argument schema, the Obligation Argument, that in fact looks capable of yielding both entitlements and obligations. According to that argument, if the conditions of concept-possession make it incoherent to doubt a practical rule's authority, whether in the guise of obligatingness or entitlingness, then that rule is in fact genuinely obligating or entitling. While there are potential problems with this argument, I hope to have shown that at least some argument in this area is promising enough to explore as a possible way of avoiding the autonomy of ethics; and so, that the argument's key premises are worth engaging.

So let us investigate those key premises. Are there concept-constituting inferential transitions that are ethically relevant? And are there concepts such that, while having them requires accepting certain inferential transitions, no would-be skeptic about those inferential transitions can lack those concepts? Examining these questions will lead us to further identify two related argument schemata, focused not on the conditions of concept-possession, but rather directly on the conditions of employing a given concept in judgments or in reasoning.

4.3 ETHICAL CONCEPTS, THEIR EMPLOYMENT, AND ETHICAL COMMITMENTS

How to argue that *no* concepts satisfy the conditions that a metasemantic vindication of an ethical view would require them to satisfy? We can of course investigate some of the prime suspects: the concept of authority, for instance, as the would-be skeptic might have to deploy it to formulate her doubts; or the concepts *ought* and *reason,* these being made pertinent by the suspicion that they are, on the one hand,

central to thinking about ethics in our broad sense, and on the other, associated with a particular disposition of reasoning: namely, the disposition to be moved in accordance with one's normative judgments (*normative judgment internalism*).[193] Yet further suspects are the concepts expressed by so-called 'thick' ethical terms, such as 'justice', 'moderation', 'honesty', 'promising': one might think that understanding these terms necessarily involves one's taking justice, honesty, etc. to be genuinely reason-giving or obligating.

The question in each case is: is there some specific practical rule R that one must be disposed to follow, or that one must take to be authoritative, in order to possess the target concept? In each case, I will argue for a negative answer. But while this piecemeal approach can eliminate the prime candidates, the suspicion remains that further concepts not yet investigated might satisfy the metasemanticist's strictures. To respond to this worry, I will do two additional things: First, I will narrow the field of prime suspects to 'thin' terms that are central to ethics, in our broad sense of 'ethics'; and second, having argued that the prime suspects fail to yield a metasemantic vindication of an ethical view, I observe that it is hard to see what central concept of ethics might be missing.

Someone could of course always attempt to define a new putative concept in relation to some inferential transitions, *stipulating* that one lacks the relevant concept unless one is disposed to engage in some specific types of inferences, or has some specific beliefs. But if one's readers or oneself can understand the target concept of such stipulations, this must then be either because one's stipulations are false, or because one's readers share the stipulated dispositions or beliefs. Yet how could one stipulate that someone has a disposition or a belief? I take it that such stipulations will not help the metasemanticist.

---

[193] One might further think that the concepts *means* and *end* are associated with a particular disposition of practical reasoning, namely a disposition of instrumental reasoning, in a way that vindicates instrumental reasoning. But I argued in chapter 2 that thought *about* means and ends is not thereby reasoning that is instrumental in form. It follows that one can have the concepts *means* and *end* without the disposition towards instrumental reasoning in particular.

So let us start with thick terms. This will allow us to eliminate them as candidates for yielding a metasemantic vindication of an ethical view, and thus to narrow the field to thin terms.

4.3.1 Thick terms and their uses

Thick ethical terms such as 'honest', 'just', 'brutal', and 'treacherous' are ethical in the narrow sense contrasted with, for example, thick aesthetic terms such as 'graceful' or 'clumsy'. Our question is whether any thick terms—whether ethical in this narrow sense, aesthetic, or something else—require one to be disposed towards some particular type of practical reasoning R, or to in some way take R, and R-based reasons, to be authoritative. What is distinctive of thick terms is that they have both a descriptive and an evaluative or normative dimension.[194] This dual aspect might suggest that thick terms, and the concepts they express, are good candidates for grounding ethical views. While thin terms such as 'ought' and 'reason' can seem, at least prima facie, *so* thin as to lack any essential link to any contentful view about how one should reason,[195] thick terms, in contrast, precisely have descriptive content that looks to forge such a link. Discussing such a possibility, Bernard Williams observes the following about the "vocabulary of promising":

> It is hard, for instance, to use the vocabulary of promising and at the same time to sustain the position that there is nothing decisive to be said, for or against, on the question whether one ought to keep promises. (1985: 26)

The suggestion is that one who employs the vocabulary of promising must hold, in some way, the position that there *is* "something decisive to be said, for or against," something in particular, namely, keeping one's promises. And such a position is, of course, a position about what practical reasons there are for what, and so an ethical view, or a part of one, in our broad sense of 'ethics'. Parallel suggestions might be made about other thick terms. The suggestion is less plausible in the case of thick aesthetic terms: while they have a descriptive as well as an evaluative aspect, their evaluative dimension seems to

[194] I will take 'evaluative' and 'normative' to be equivalent here.
[195] Whether thin terms really are, in this way, as thin as they seem to be is the topic of §4.3.2.

lack any essential action-guiding aspect. But the evaluative dimension of thick ethical terms, in the sense of 'ethics' contrasting with aesthetics, is often thought to have such an action-guiding aspect. A prospective action's justice looks to be not (only) a reason to passively delight in that aspect of the action, should the action be performed, but rather, a reason to actively choose to perform it.

However, while it is plausible that how one uses certain terms in context can reveal (parts of) one's take on the question how to live—our uses of words can reveal our normative commitments—this does not yet help the metasemanticist. For in this regard, pejorative terms such as 'boche' are no different from thick ethical terms. Ordinary (i.e. pejorative) uses of 'boche' plausibly reveal some of the user's normative commitments regarding what is a good reason to think or do what; but as we saw, it does not follow that, to possess the concept *boche*, one must engage in those ordinary uses. Instead, one who lacks the normative commitments implied by ordinary uses of 'boche' is likely to refrain from those uses precisely because she lacks the relevant commitments, and yet understands 'boche' well enough to realize what those uses would communicate. Similarly, ordinary discourse concerning promising or justice may well reveal that the speaker thinks there is something to be said for promise-keeping or just actions; but that does not prevent someone else from refusing to (sincerely) partake in such discourse, precisely *because* she well understands the meaning of, and the normative commitments undertaken in, using the "vocabulary of promising" or of justice.

So even if the ordinary use of words entails that the speaker has certain normative commitments, this does not in turn entail that possessing semantic understanding of those words entails one's undertaking those normative commitments. On this point, Williams agrees:

> An insightful observer can indeed come to understand and anticipate the use of [a] concept without actually sharing the values of the people who use it […] [I]n imaginatively anticipating the use of the concept, the observer also has to grasp imaginatively its evaluative point […] But […] he may not be ultimately identified with the use of the concept: it may not really be his. (1985: 141-142)[196]

[196] See also Williams 1985: 26. Compare, in a different context, McDowell 1981, §2. McDowell thinks that (at least) imaginative appreciation of the evaluative point of a thick concept might be required for one to even have a grasp of its *descriptive* dimension; he therefore protests against assuming that a "factorizing" analysis of thick terms, into separately intelligible evaluative and descriptive dimensions, is possible. I touch upon the "factorizing" issue below.

The same seems to be true of thick terms in general. If certain uses of such terms entail normative commitments, then one who understands the terms can studiously avoid those uses. One way of exhibiting semantic understanding is precisely such avoidance, at least if the avoider can explain her avoidance by mentioning the term in question and using other words. The Obligation Argument, and indeed any metasemantic argument, cannot get going with thick concepts, as its first premise is false.

However, might a metasemanticist retreat to an argument on which it is not a concept's *possession* conditions, but instead just the conditions of its *employment* in judgments or in reasoning, that vindicate an ethical view? Such an argument may more properly be called *metapragmatic* rather than metasemantic. Of course, as we saw, the metasemanticist views the possession and use of concepts as closely related: there are, according to her, certain privileged uses that just *are* conditions of concept-possession. However, even if we reject that thought, as we just did as regards thick concepts, we might nonetheless attempt the following argument:

The Use Argument for Practical Obligations

(11U) Anyone using the concept C must in some way take RO to be genuinely obligating, at least implicitly.
(12U) Any would-be skeptic about RO must use C: one could not even formulate one's doubts about RO without using C.
(13U) So, any would-be skeptic about RO must take RO to be genuinely obligating, at least implicitly. (From 11U and 12U)
(14U) So, it is incoherent to doubt RO's being genuinely obligating: skepticism about RO is self-defeating. (A restatement of 13U)
(15U) If it is incoherent to doubt that *p*, then *p*.
(16U) So, RO is in fact a genuinely obligating rule.

This argument will not help, however. The chief problem with it stems from the fact that 'use' is ambiguous between various possible types of use. In particular, there are *distancing* uses that either implicitly or explicitly cancel the usual normative implications of non-distancing uses. And while (11U) requires us to read 'use' as referring to non-distancing uses, (12U) requires us to read 'use' as referring to

distancing uses. To make their premises plausible, the Use Arguments are forced to equivocate. Let me explain.

What is a distancing use? One example is a *quoting* use, as e.g. when I report someone's utterance by saying "Paul said, 'George promised'." Some may wish to count this as a case of mention, not use. But this sort of case has the peculiarity that the quotation is a device for communicating *what Paul said*—i.e. *that George promised*—not only the words Paul used to say it; and this suggests that what is in question is a special quoting use of words. (We might think of this as a use that simultaneously manages to be a mention.)[197] What is interesting for our purposes is that this type of quoting use does not carry the same normative implications as does, e.g., my use of the concept of promising when I exclaim, "Paul said that George promised!" While the exclamation does seem to communicate that I have, or undertake, the normative commitments that we are assuming ordinary, non-distancing uses of the "vocabulary or promising" to import, the quotation does not: it distances the speaker from those ordinary normative implications. In quoting, a speaker may grasp what the normative commitments undertaken or revealed would be in an ordinary, non-distancing use. But because the speaker's use is distancing, those commitments are not, as Williams might say, his. (Compare, in this same regard, the two utterances "Paul said, 'George is boche'" and "Paul said that George is boche!" The first may be a use available even to one who wishes to studiously avoid the normative implications of ordinary uses of 'boche'.)

How does the distinction between distancing and non-distancing uses help to show that the Use Argument equivocates? Here is how. To make (11U) plausible, 'use' must be read as referring to non-distancing uses of C, the uses that, we are supposing, do commit one to some practical norm R. But on that same reading of 'use', (12U) is just false. For the skeptic can, and plausibly even must, express her skepticism about RO by means of linguistic distancing devices akin to quotation—devices by means of which one can signal that one disagrees with, or remains non-committal with regard to, the normative commitments that some non-distancing uses import. Which linguistic distancing devices may the skeptic

[197] For a very helpful overview of controversies surrounding quotation, see Cappelen & LePore's 2005/2009 *Stanford Encyclopedia* article "Quotation."

use? One type of device might incorporate quotation, e.g. "I doubt that the rule 'RO' [e.g. 'Keep your promises!'] is authoritative." But more simply, a skeptic about the ethical import of e.g. considerations of justice might say "I doubt that an option's justness makes it choiceworthy." Such devices serve to *explicitly* cancel the putative ordinary evaluative implication of the terms used. In general, then, the problem is that while the skeptic is likely to formulate her doubts by means of distancing uses, such distancing uses are precisely designed to cancel the normative implications present in the types of use that stand a chance of making (11U) true. In general, all thick terms seem to be susceptible to such distancing devices; and so their putative ethical import is coherently subject to skepticism.[198]

One might object to my supposition that the skeptic might explicitly distance herself from the putatively ordinary evaluative implications of some thick ethical term by using, in her own voice, *thin* terms, to the effect that an option's justice is not always a reason, or a decisive reason, in favor of it. For one might hold that the ordinary evaluative import of thick vocabulary is not in fact factorizable-out from its descriptive import in a way that would allow us to cancel that evaluative import by using thin terms.[199] However, even if this opposition to a "factorizing" view of thick terms is correct, it will not help in the context of attempting to provide a non-question-begging argument for an ethical view. If the normative import of thick terms is somehow different from the normative import of thin terms like 'reason' and 'ought', then that normative import does not automatically speak to Socrates' question, understood as a

---

[198] Of course, not all attempts at distancing work: one who says "She is boche, but there is nothing wrong with that" is not very convincing. However, such attempts at explicit cancellation do seem to work in the case of e.g. the above utterance about justice. It is an interesting question why there aren't as many distancing uses available for pejoratives as there are for thick ethical terms. Here is one possible explanation. The assertion "an option's justice does not make it choiceworthy" commits the speaker to the claim that there *is* such a thing as an option's justice, and thereby to the claim that the concept *justice* finds real application in the world; and likewise the assertion "a person's being boche does not make that person bad" commits the speaker to there being such a thing as someone's being boche, and thereby also to the claim that the concept *boche* finds real application in the world. But one who regards ordinary uses of 'boche' as objectionable is likely to think that the concept *boche* does *not* find any real application in the world; therefore she will also refuse to assert claims like "a person's being boche does not make that person bad." That claim would not function as a relevantly distancing use of 'boche' for such a one. In contrast, one who doubts that considerations of justice have genuine ethical import need not thereby think that the concept *justice* lacks real application in the world; all that such a one doubts is whether an option's (actual) justice speaks in favor of it. Assertions like "an option's justice does not speak in favor of it" do therefore count as relevantly distancing uses for such a one.

[199] In a different context, McDowell argues against the assumption of factorizability in his 1979, 1981. I do not mean to suggest that McDowell would object to the point I go on to make.

question about good practical reasoning and reasons. For that question is formulated in thin terms: in terms of what good practical reasoning and reasons are. On the other hand, to hold that the putatively unfactorizable-out normative import of e.g. justice *does* speak to Socrates' question is just to hold that the specific way in which justice purports to count in favor of an action, a way that is putatively uncapturable in thin terms, is a way in which we should think of considerations as genuinely counting in favor of actions. And this is just to stake out an ethical view: it is a view about the distinctive normative perspective, or one of them, from which questions about reasons for action and good reasoning are properly answered. So even on the non-factorizing view, thick terms cannot help us avoid the autonomy of ethics.

Nor would it, finally, help to suggest a different Use Argument, one for Entitlements:

The Use Argument for Practical Entitlements

(1U) To use the concept C, one must be disposed to reason in accord with R.
(2U) One is always at least defeasibly practically entitled in using any concept.
(3U) If one is at least defeasibly practically entitled in using a concept, then one is (thereby) at least defeasibly practically entitled in satisfying the conditions of using that concept.
(4U) So, one is always at least defeasibly practically entitled in satisfying the conditions of using C.
(5U) So, one is always at least defeasibly practically entitled in being disposed to reason in accord with R.

For if we restrict 'use' in (1U) so as to exclude ethically non-committal uses such as quoting uses, we have to interpret (2U) as involving that same ethically committal sense of 'use'; but precisely because that sense of 'use' is designed to include only ethically committal uses, this makes (2U) itself ethically committal in a way that makes the argument question-begging. (Recall, as before, that practical entitlement to a use must be (practical) entitlement to a *practical* use; and to assume that a given practical use is one to which the standards of practical reason entitle us is simply to assume a first-order claim about ethics.) (2U) might even be *true* in some ethically committal sense of 'use', but it cannot help us avoid the autonomy of ethics.

I have focused on one central type of problem with the Use Arguments; that there are distancing and non-distancing uses of thick ethical terms, and that the Use Arguments must equivocate between the two in order to make their premises individually plausible. There are, of course, other problems with the Use Arguments as well. For instance, what the normative commitments imported by so-called "ordinary" uses of a thick ethical term are is itself a controversial question. Many thick terms admit of seemingly opposed normatively committal uses, depending on context. Some people (and I hope most) deplore treachery and cruelty, and regard an action's having these characteristics as a strike, even a decisive one, against it; but it is also perfectly intelligible that someone should view such features as particularly delicious aspects of an action, taking and savoring opportunities for treachery or cruelty wherever possible. Such a person perhaps *need* not take herself to be living as she should; but the point is that she might, for all that the concepts of treachery and cruelty tell us.[200] Do all thick terms admit of opposing views of their ethical import? Certainly it has proved hard to disprove Thrasymachus' view that justice is not a virtue of practical reason, something that would be quite surprising if the view were in fact nonsensical. And it is hard to see why justice should be special in admitting of opposing views of its ethical import. If the normative valence of thick terms can shift in this way depending on one's pattern of use, then selecting for attention just one such use among many would be question-begging.[201]

I will not dwell further on these problems here. Whatever their fate, if my main argument stands, then thick ethical concepts cannot help us vindicate an ethical view. Neither the conditions of their possession nor the conditions of their use can vindicate either entitlements or obligations.

What about thin concepts? In the normative sense of 'reason', the term 'reason' does not seem capable of switching ethical valences depending on context of use. Likewise for terms such as 'ought', 'authoritative', 'practically rational', 'good practical reasoning', and 'virtue of practical reason'. These, or

---

[200] Compare Williams 1985: 91.

[201] Problematic ambiguities likewise afflict the *descriptive* dimension of thick terms. There is often a longish distance between, on the one hand, what a putatively "ordinary" use of a thick *concept* commits one to, and on the other, what a somewhat different pattern of use evincing a particular *conception* of the thick ethical characteristic in question commits one to; and selecting a particular use as the one to focus on may already involve, question-beggingly, selecting a conception as the ethically right one.

other terms expressing the same concepts, *just are* the central terms in which any ethical view deals. However, on the face of it, it also seems that the thinner the concept, the less likely it is to have conditions of either possession or use that commit one to some specific ethical view over another. For while these concepts have a settled normative dimension, their thinness consists, precisely, in their lacking a determinate descriptive dimension. It is hard to see how merely having or using them could commit one to any contentful ethical view over another. They are, instead, simply the concepts in terms of which *any* ethical view, no matter how strongly conflicting with other views, may be stated. Can argument overturn this prima facie impression?

4.3.2 Thin terms and their uses

I will be discussing the conditions of use for thin terms. If the conditions of using thin terms are ethically non-committal, then so are, *ipso facto*, the conditions of understanding them. And I will be arguing that the conditions of use for thin terms are ethically non-committal.

Now, there are of course many ethically committal uses of thin terms, just as there are ethically committal uses of thick terms: for example, the sincere assertion that a certain practical rule R is authoritative is an ethically committal use of 'authoritative', since this use is possible only if the speaker believes, and is in this sense committed to, what he sincerely asserts. But as before, the relevant question is not whether there are some ethically committal uses, but rather, whether there is an appropriately non-question-begging and unequivocal sense of 'use' in which the use of thin terms is ethically committal. To this question, I will argue that the answer is no.

It will be helpful to start by examining whether our earlier distinction between distancing, i.e. ethically non-committal, and non-distancing, i.e. ethically committal, uses can be employed in an argument parallel to the one I gave above in the case of thick terms. Can we argue, directly, that any instance of each Use Argument, where C is a thin concept, must equivocate on 'use' in order for its

crucial premises to be non-question-beggingly true? It turns out that, with one of the Use Arguments, we cannot. Matters will instead be somewhat more complicated here than in the case of thick terms.

4.3.2.1 The Use Arguments and the charge of equivocation

Take first the Use Argument for Practical Obligations. In the case of thick terms, recall that the skeptic in premise (12U) can express her skepticism about, say, the reason-givingness of considerations such as "I promised" by distancing, ethically non-committal, uses of the relevant thick vocabulary. This created trouble for the Use Argument for Obligations because it is crucial for the success of the argument not only that there is some non-distancing use that renders an instance of (11U) true, but also that there is no available distancing use sufficient for formulating the doubts in (12U). In the case of thick terms, a distancing use that is sufficient for the skeptic's needs is always available.

However, there are obstacles to running exactly the same argument in the case of thin terms. For the skeptic's "distancing" in the case of thick ethical terms was achieved by her use—ordinary, non-distancing use—of thin ethical terms. To express her doubts about the genuine reason-givingness of considerations of promising, it seems that the skeptic *has* to say something to the effect of "I doubt that having promised to φ is really a good reason for one to φ"; or "I doubt that one should keep one's promises"; or "I doubt that the rule 'Keep your promises!' is authoritative." We can state the problem more generally. Since the skeptic is a skeptic *about* the genuine ethical import of some putative practical rule R, and since such doubts must always be expressed by using thin ethical vocabulary, the skeptic cannot, in formulating her skepticism, distance herself from that thin ethical vocabulary—at least in its normative dimension: she must use it in its ordinary normative signification.

It does not help to suggest that the skeptic might choose *different* thin vocabulary in which to express her doubts than the vocabulary that expresses the concept C whose conditions of use are in question in the argument. For all the thin ethical vocabulary in which a skeptic might express her doubts seems to be analytically connected in some way. 'Reason' and 'ought' seem trivially connected in that one ought to do what one has decisive reason to do; and while the connections between the meaning of

'rational' on the one hand, and the meaning of 'reason' and 'ought' on the other, are disputed, there does seem to be *some* connection, if 'rationality' is relevant to issues of good practical reasoning and reasons at all—whether it is that reasons are to be defined in terms of practical rationality, vice versa, or some more subtle connection. 'Authoritative' is relevant only because it signals something like 'genuinely reason-giving' or 'genuinely obligating' ('obligating' being related to 'ought'). We already noted the relation between 'obligated', 'permitted' (or 'entitled'), and 'forbidden'. 'Good practical reasoning', in turn, signifies good ways of moving from considerations to actions, such that those considerations (i) thereby become the reasons for which one acts, and (ii) are also good reasons to act on. Anyone who thinks that people's reasons for acting can sometimes be good ones thereby also thinks that people's practical reasoning can sometimes be good.[202] And finally, 'virtue of practical reason' just signifies a good disposition of practical reasoning. Given all of this, it seems that a skeptic could not express her doubts about the genuine ethical import of some practical rule R in terms of one thin term without thereby committing herself to the denial of the ethical import of R in terms of all the others.

More strongly, if the skeptic wishes to express the position that a putative practical rule R has no genuine ethical import whatsoever, she is therefore trying to express the position that R is not genuinely reason-giving, not obligating nor permitting, not authoritative; that following R does not constitute good practical reasoning, nor is it a way of being practically rational; and that the disposition to follow R is therefore not a virtue of practical reason. Since the skeptic about R's genuine ethical import is committed to all of this—or if I have left some important thin concept out, also to the denial of that concept's correctly applying to R—there are simply no thin ethical concepts left over that the skeptic might refuse to employ in their ordinary normative signification consistently with her skepticism about R's ethical import.

---

[202] There are of course more worked-up notions of practical reasoning as well. But since more worked-up, they are also more controversial. For some defense of the present, permissive notion of practical reasoning, see my chapter 3, §3.2. At any rate, the thin notion of practical reasoning made use of here is parallel to the way Boghossian thinks of theoretical reasoning: any step from a putative reason to believe *p* to the belief that *p* counts as an inferential step, a step whose goodness an epistemic view might assert or deny (see e.g. 2008a).

If a distancing use of some thin ethical term C is possible for the skeptic, then, it will have to be a use distancing her from some putative *descriptive* dimension of C—from some ethical commitment that ordinary non-skeptical uses of C putatively entail. Now, our prima facie impression was that ordinary uses of thin terms lack such determinate ethical commitments: they can instead be used to state any among a variety of competing ethical views. Of course, as we noted, there are many ethically committal uses. But it is also easy to find ethically non-committal uses. For example, Paul's assertion "George claimed that Sarah ought to φ in *c*" does not commit Paul to any ethical view; Paul merely uses 'ought' to communicate that George claimed something ethically committal. The crucial issue is not whether there are ethically committal or non-committal uses—there clearly are—but rather, whether there is any use of thin terms that is *both* ethically committal *and* one in which the would-be skeptic cannot avoid engaging.

The example of sincerely asserting an ethical view, though clearly an ethically committal use, seems to fare badly in this regard. Given any assertion to the effect that some rule R is authoritative, it seems that a skeptic about R's authority precisely does *not* engage in that use; she is precisely concerned to doubt, not assert, R's authority. That, at any rate, is the flat-footed response to such examples. A similarly flat-footed response seems to be available to more subtle proposals about the ethically committal use that makes (11U) true. Suppose that the ethically committal use in question is not a matter of sincere assertion and attendant belief, but rather, a more subtle matter of one's somehow "taking" the rule R one is following to be genuinely obligating, where such "taking" is not just belief. Still, why must the skeptic of (12U) engage in *that* use, if she is precisely concerned to doubt that R is genuinely obligating?[203] Can she not explicitly distance herself from that use by saying something to that effect? The challenge for the Use Argument for Obligations is to propose an ethically committal use of some thin ethical term from which the skeptic cannot coherently extricate herself by simply denying the very ethical commitment in question. This challenge clearly falls short of an argument that any instance of the Use Argument for

[203] One possible answer would claim that the skeptic must take RO to be obligating because she is an agent and every agent just as such must take RO to be obligating. But again, I have argued that the nature of agency is normatively neutral in chapters 2-3. So this avenue will not work.

Obligations, where C is a thin concept, *must* equivocate on 'use' in order for its crucial premises to be true. But it does show that, unless we can find some use of a thin term that the skeptic must engage in even while that same use (probably implicitly) commits her to the very ethical view she wishes to doubt, the Use Argument for Obligations fails.

Things seem to be more straightforward in the case of the Use Argument for Practical Entitlements. As before, it is easy to find an ethically committal, restricted sense of 'use' on which the use of some thin concept C makes (1U) true: for instance, the use of *ought*, by someone who is disposed to reason in accord with the rule "Whenever your mother says you ought to φ, φ," in reasoning in accord with that rule. The difficulty is in holding that, in that same restricted sense of 'use', one is practically entitled to use *ought*, as per (2U), *without* thereby merely presupposing that one is entitled to a certain type of practical reasoning. Since (2U) just states a putative practical entitlement, it is hard to see how we could avoid begging ethical questions in asserting it.

From here on in, then, I only consider the Use Argument for Obligations. Is there some use of a thin term that the skeptic of (12U) must engage in even while that same use (implicitly) commits her to the very ethical view she wishes to doubt?

4.3.2.2 Normative judgment internalism, and beyond: the coherence of skepticism

There is, in fact, one type of employment of thin normative concepts that may be thought to fulfill the criteria outlined for the Use Argument for Obligations to succeed. This is their employment in 1st-personal judgments about what one ought to do, or what one has most or decisive reason to do. Several philosophers have proposed that such judgments are essentially linked to a particular practical disposition, namely, the disposition to be moved, by those very judgments, to act in the way that one judges one ought.[204] The idea may be summed up in the slogan "normative judgments necessarily motivate"—an

[204] For defenses of either this general internalism about normative judgments, or of internalism about moral judgments more narrowly, see e.g. Wedgwood 2007: 25, 27-8, 153-173; Smith 1994: 69-76; Pettit & Smith 1996; Burge 1998: 251-2. For criticisms, see e.g. Arpaly 2003: ch.2, Svavarsdóttir 1999, Schroeter 2005.

idea that has received various different formulations, all gathered under the moniker *normative judgment internalism*, or **NJI** for short. How might some version of NJI help in discovering a sound instance of the Use Argument for Obligations? Consider the following argument, which I will call the 'NJI Argument'.

Suppose that 1st-person normative judgments necessarily dispose one to be moved by those judgments to do as one judges one ought. Let us say that, in making such judgments, one is necessarily disposed to reason in accord with (something like) the rule

NJI-R If you judge you ought to φ, φ!

If anyone making 1st-person normative judgments must be disposed to reason in accord with NJI-R, does this make an instance of (11U) true? We might think it does not, because of the general point that we have already made: that having a disposition to reason in accord with a rule looks compatible with doubting the rule's authority. But NJI-R looks to be special. Suppose one does make a 1st-personal judgment "I ought to φ"; and suppose that NJI is true, so that in making such a judgment, one is thereby disposed to φ. Might someone who is so disposed, and who does make a judgment "I ought to φ," nonetheless doubt that NJI-R is authoritative? It seems that to doubt NJI-R's authority would be, in effect, to doubt whether one should do what NJI-R tells one to do in one's situation. But what NJI-R tells one to do is to φ. And that is just what one judged one ought to do. So if one is to maintain one's first-personal judgment that one ought to φ, then one cannot simultaneously doubt that one ought to φ. This means that one cannot simultaneously make a first-personal judgment that one ought to φ and doubt the authority of NJI-R—the very rule that one must be disposed to follow in making 1st-personal judgments at all. On this reasoning, anyone engaging in a certain use of 'ought', namely its use in judgments of the form "I ought to φ," does implicitly commit herself to taking NJI-R to be genuinely authoritative. So the relevant instance of (11U) is true.

In fact, this argument does not even depend on the idea that NJI is true. One might lack the motivational disposition, and still be implicitly committed to the authority of NJI-R merely in making judgments of the form "I ought to φ." All that the argument really depends upon is the idea that one

cannot simultaneously use 'ought' in sincere judgments of the form "I ought to φ" and doubt that one ought to do what NJI-R would tell one to do in the situation, namely, to φ. We could run the same argument with the concept *reason*. One cannot simultaneously judge that one has reason to φ, and yet doubt that one has reason to do whatever is correct by the lights of a rule enjoining one to do what one judges one has reason to do.

Even if this is correct, of course, it does not yet establish NJI-R's authority: it only establishes that one cannot doubt NJI-R's authority *while making 1st-personal ought-judgments*. But if we could further argue that a would-be skeptic about NJI-R's authority must make 1st-person ought-judgments in framing her skepticism, then the crucial premises of the Use Argument for Obligations would be shored up. The skeptic would have to engage in a use of a normative concept, namely, its use in 1st-person ought-judgments, that implicitly commits her to the authority of the very rule that she purports to doubt. We would, at least, have vindicated one objective practical norm without question-begging: namely, that one ought to act as one thinks one ought. This completes the NJI Argument.

So is there room for skepticism about NJI-R's authority? It would certainly be startling if there were not. For we ordinarily think that we can be wrong in our judgments about what we ought to do—something that the conclusion that one ought to do what one thinks one ought to do leaves no room for. Now, we have already seen that a skeptic about a rule's authority must make a normative judgment of *some* sort using thin terms. And the NJI Argument seems to show that one cannot simultaneously doubt NJI-R's authority and make 1st-personal judgments of the form "I ought to φ." So it looks like the only avenue for skepticism about NJI-R is to deny that the skeptic must make 1st-personal ought-judgments in formulating her doubts. Can she formulate her doubts in (12U) without engaging in the use that renders the relevant version of (11U) true?

The most natural way for a skeptic to express her doubts about NJI-R's authority or obligatingness seems to be in the form of some general claim like "I doubt that NJI-R is authoritative," or "I doubt that I always ought to do what I judge I ought." Neither of these is, or incorporates, a judgment of the form *I ought to φ*. But one might think that the skeptic's general judgment at least implies that she

must also be making such a 1st-person judgment. For her doubts are about whether NJI-R is a good rule of reasoning for agents just as such—about whether following NJI-R is part of how agents just as such should live. If she herself is an agent, or more importantly, believes that she is, then won't her judgment that agents in general ought not to follow NJI-R imply that she is also judging, at least implicitly, that she herself ought not to do so? And isn't this a 1st-person judgment of the relevant sort?

In response, first, skeptics might or might not believe that they are among the people to whom the rules whose authority they doubt applies. But second, and more important, a skeptic's doubts about the authority of a rule need not take the form of believing that one ought *not* to follow that rule. They might instead take the form of a suspension of judgment: "I doubt that NJI-R is authoritative" need not imply that one is quite sure that NJI-R is not authoritative, only that one is quite unsure that it is. Thus the skeptic need not judge "I ought not to follow NJI-R"; she can simply judge "I doubt that NJI-R is authoritative." So it seems that there is a way for skeptics to formulate their doubts about NJI-R without thereby implicitly committing themselves to NJI-R's authority.

A defender of the NJI Argument might protest that the argument at least shows that as soon as one *does* make 1st-person judgments of the form "I ought to φ," one commits oneself to NJI-R's authority. And surely most of us have already made such judgments. So we are all committed. Never mind that NJI-R can be doubted by some arbitrary skeptic; at least it cannot be doubted by anyone who has any ethical views at all about how they themselves ought to act. All such people, at any rate, ought to do whatever they judge they ought to do.

In fact, however, this protest reveals a deeper confusion in the NJI Argument. In formulating that argument, we were wrong to say that "to doubt NJI-R's authority would be, in effect, to doubt whether one should do what NJI-R tells one to do in one's situation." If one has already judged that one ought to φ in one's situation, then *of course* one cannot simultaneously doubt that NJI-R, which in one's situation gives one the verdict that one ought to φ, yields, in that situation, a correct verdict. But this is compatible with doubting the authority of NJI-R more generally. For we can think that we are fallible in our judgments about what we ought to do. (Indeed, it is surely a plausible assumption that we are in fact

fallible.) And one's recognition of one's fallibility can lead one to judge, justifiably it seems, something of the form "I doubt that I always ought to do what I judge I ought [and I doubt this because I think my ought-judgments may often be false]." This would be a type of distancing use of 'ought'—a use distancing the skeptic from the putative normative commitments imported by making judgments of the form "I ought to φ." The rationale for the doubt expressed in the distancing use is just the perfectly ordinary recognition of fallibility: the recognition that sometimes our judgments about what we ought to do lead us astray if we act on them. This is a doubt that each of us can coherently entertain; its possibility is not confined to an implausibly abstract skeptical figure.

So even though any actual judgment "I ought to φ" commits one to thinking that NJI-R's verdict about what one ought to do given that judgment is true, it does not follow that one cannot doubt NJI-R's authority. One can acknowledge that NJI-R's verdicts coincide with the truth about what one ought to do in the cases where one's initial ought-judgments are themselves true. But this does not commit one to thinking that NJI-R's verdicts are in general authoritative, or true. For one can realize that (or at any rate believe that) many of one's initial ought-judgments might be false.[205]

This allows us to make a more general point, concerning all thin terms. Observe that the NJI Argument traded on a general feature of NJI-R to try to support it: that one's judgment "I ought to φ" is both what gives NJI-R application in one's situation, and simultaneously commits one to thinking that the verdict that NJI-R yields in one's situation is correct. Now, the general point I wish to make in the light of this observation is two-fold.

First, it seems that we can invent a rule with the same feature for any thin term, and thereby "trap" anyone who follows the rule into thinking that the verdicts it yields, as one follows it, must be authoritative. I already mentioned the concept *reason*. Suppose that in the course of following some rule

---

[205] A proponent of the NJI Argument might object that, if NJI-R is in fact an authoritative rule, then one's initial ought-judgments *cannot* be false: the fact that NJI-R is authoritative makes them true. But this response on behalf of the NJI Argument would not help in attempting to get beyond the autonomy of ethics. It relies on the first-order normative claim that NJI-R is in fact authoritative. That is just the sort of claim that metasemantic and metapragmatic arguments are attempting to vindicate, not help themselves to.

of reasoning with this "trapping" structure, one must make sincere judgments of the form "*p* is a good reason to φ." And suppose that the rule one is following yields the verdict that, in one's circumstances, one should φ because *p*. Then one cannot, *while* following the rule, also think that the rule is in this very instance leading one astray. For in doing what the rule tells one to do, one is doing φ because *p*, which conforms to one's initial judgment that *p* is a good reason for one to φ. Nonetheless, just as with NJI-R, one can doubt that the rule one is following always and generally leads one to act for good reasons; for one can recognize one's fallibility as a judge of what is a good reason for what. Likewise for the concept *rational*. If, in the course of following some rule, I must sincerely judge that φ-ing is rational, and the rule tells me to φ in the instance, then I cannot, while following the rule, think that it would be irrational for me to do what the rule tells me to do in the instance. Nonetheless, since I can think that I am fallible, I can doubt more generally whether my judgments about what it is rational to do are correct, and thus I can doubt whether conforming to the rule's dictates is in general what it is rational to do.

One can run through the same points for the concepts *obligation*, *entitlement*, *authority*, etc. (Try it.) If it is a central concept of ethics, a concept in terms of which we might frame an ethical view, then we can always formulate a putative rule of reasoning such that it is constitutive of following that rule that, in doing so, one employs the relevant thin concept in a judgment whose content conforms to the verdicts of that very rule. But even if we thus "trap" anyone who might be genuinely following the rule in question into momentary inability to doubt the rule's present verdicts, skepticism about the rule's authority more generally still makes good sense. It is hard to see what central thin concept might be missing, or might escape this fate.

The second general point I wish to make here is that this "trapping" structure for rules of reasoning seems to me to be the only promising way to attempt to generate a sound instance of the Use Argument for Obligations. For if nothing in the course of one's reasoning commits one to any claims about the rule's goodness or about the correctness of its verdicts, even in the instance, then it is hard to see how we might motivate a version of (11U). It is easy to doubt the authority of rules of reasoning that

do not have the "trapping structure," since one might continue to follow such rules without thereby committing oneself to any views about the goodness of even their present verdicts.

Given these two points, however, we can always coherently doubt the authority of any putative practical norm, whether it has a "trapping" structure or not. This is so even if we are in fact disposed to follow the rule we doubt. For even if, in any instance of following it, one cannot doubt the goodness of its verdicts, as is the case with rules with a "trapping" structure, one can nonetheless doubt quite generally whether one's judgments and dispositions of reasoning are good ones.

Eventually, of course, such doubts might lead to problems. For suppose one does have severe doubts about the authority of a rule, say, NJI-R. Yet in being disposed to follow that rule, one is probably also disposed to make sincere judgments about what one ought to do, judgments that NJI-R in turn claims are true. Then it looks like one's skepticism and one's disposition to follow NJI-R might be at odds. In such a case, one's disposition to follow NJI-R might erode as a consequence of one's skepticism. One might even stop making 1st-personal ought-judgments altogether. Or it could happen the other way around: one's individual judgments might seem more trustworthy to one than one's general skepticism, and the skepticism might be what gives way. These are problems one might encounter in doubting a rule with a "trapping" structure that one is disposed to follow. But these are not problems concerning the very idea of doubting the rule in question. They are just problems of life that can confront one in thinking about whether one's practical dispositions are good.

My argument concerning NJI-R has, then, yielded the general conclusion we wanted. Given any use of thin terms that would commit one to some practical norm, a skeptic about that practical norm need not engage in that use in formulating her doubts. She may engage in that use *qua* person; but what is important is that she need not do so to formulate her doubts. Strikingly, the argument for this conclusion, which generalizes from the possibility of doubting NJI-R's authority, does not rely on arguing against NJI itself. For all I have said here, 1st-person normative judgments may be essentially motivating for those who make them. Nonetheless, one can doubt the general claim that one should always act as one judges one ought. As we saw, parallel points apply to other putative practical rules. If this is right, then given our

earlier arguments with regard to thick terms, skepticism about the authority of any putative practical norm is coherent.

This concludes my main argument that metasemantic and metapragmatic arguments for ethical views fail. Since one can use any thick or thin ethical term without thereby implicitly committing oneself to a view about which practical rules are authoritative—an ethical view—one can also possess any thick or thin ethical concept without thereby committing oneself to an ethical view. Neither the conditions of concept-possession, nor those of concept-use, help to yield the hoped-for vindication of an ethical view in normatively non-question-begging terms.

I close §4.3 by considering an objection to the foregoing. §4.4 concludes this chapter, and this dissertation, by completing our argument for the autonomy of ethics, and by briefly considering its consequences.

4.3.3 An objection: Wedgwood and the constitutive ideal of rationality

One might object that the conclusion that metasemantic and metapragmatic arguments fail is premature. In particular, one might think that, in structuring our most recent discussion in terms of the Use Argument for Obligations, we have made it too easy to refute metasemantic and metapragmatic arguments. For it is enough, in general, to invalidate the Use Argument for Obligations, that one finds appropriate distancing uses of the normative concepts whose use is in question in the argument. A stronger argument, the objection runs, would rely not on the possibility of such deviant uses, but rather on showing that there are not even general dispositions towards a type of ordinary use that could vindicate ethical views. Of course, as we saw, the Use Argument for Entitlements did rely on the idea of such general dispositions, not on the possibility of individual skeptical uses. And that Argument failed to support an ethical view without question-begging, since one of its premises must presuppose entitlement to a practical use of a concept. Nonetheless, we might think that there is a different, more promising line of argument available that

likewise relies on the idea of dispositions to use a concept in reasoning. This argument is Ralph Wedgwood's, in his (2007) *The Nature of Normativity*.

Wedgwood argues on perfectly general grounds that (a) the possession conditions of any concept, including any normative concept, must be a matter of the individual's dispositions to use that concept in reasoning in certain ways. For instance, Wedgwood thinks that the basic disposition of reasoning one must have in order to have the concept *ought* is just a version of NJI-R.[206] And he argues, moreover, on perfectly general grounds that (b) the dispositions in which each individual's possession of a concept consists must be one's rational dispositions, not one's irrational ones.[207] Notably, the argument for (b), in particular, does not depend upon our presupposing any specific claims about the content of rationality, contrary to what we might have thought; so it looks not to attract the objection that we are begging ethical questions in the course of the argument. If Wedgwood's (b) is true, then any dispositions built into the possession conditions of a concept are indeed rational ones; so if we could only discover what those dispositions are, perhaps a metasemantic vindication of an ethical view would be possible after all.

Interestingly, Wedgwood himself does not try to avoid the autonomy of ethics. His argument for the claim that a disposition to reason in accord with NJI-R is constitutive of possessing the concept *ought* is *premised* on the claim that NJI-R-following is rational, in the commending sense of 'rational' contrasting with 'irrational'. The metasemantic claim is supposed to be a consequence of this claim about rationality, together with (b).[208] In this way, Wedgwood's claims about the conditions of concept-possession are an exercise in transcendental psychology: given only very general considerations about what concept-possession must be like, the normative facts are supposed to allow us to deduce specific psychological claims.

Nonetheless, we might have hopes for a reversed procedure. So how does Wedgwood argue for the general claims (a) and (b)? Wedgwood's arguments for (a) and (b) are very general. Here is his

---

[206] See Wedgwood 2007: 25, 27-8, 153-173.
[207] 2007: 168-9.
[208] See Wedgwood 2007: 25, 27-8, 153-173.

argument for (a). Suppose there are two different communities, A and B, the members of each of which possess some concept C or C*, respectively. And now suppose that the members of community A use C in all of their reasoning exactly like the members of community B use C* in their reasoning. Then it seems absurd to suppose that C and C* might be different concepts, such as the concepts *cat* and *bicycle*. So, Wedgwood concludes, it must be the case that what determines which concept one possesses is just one's dispositions of using the concept in one's reasoning (together perhaps with facts about one's environment and the dispositions of other members of one's community) (2007: 166). If this argument is sound, then something like (a) is true (though the parenthetical remark about facts outside the individual is potentially important).

To argue for (b), then, Wedgwood reasons as follows. It is hard to see how a thinker could possess C only in virtue of her irrational dispositions, such as dispositions towards fallacious reasoning. For instance, supposing C is the concept *conditional*, it seems that one could not have it if one were only disposed towards fallacious types of reasoning such as affirming the consequent. But it is easy to see how a thinker could possess C only in virtue of her rational dispositions. For surely a perfectly rational being could possess all of the concepts there are to possess, except perhaps for any defective concepts there might be. Furthermore, Wedgwood claims, if a thinker's possession of a concept did rest on his irrational dispositions even in part, then "any use that the thinker makes of this concept would depend on his irrationality"; and Wedgwood doubts whether this could be the case, since possessing a concept is a "cognitive *power* or *ability*—not a cognitive defect or liability" (2007: 168-9). He concludes that if one possesses a concept C, this must be due to one's rational dispositions of reasoning, not due to one's irrational ones (2007: 169). So (b) is true.

What to make of this argument? The argument for (b) is problematic in its claim that an individual thinker's concept-possession could not rest on his irrational dispositions, whether in part or whole. This claim is needed if (b) is to help us sort the good from the bad types of reasoning, given a putative specification of the conditions of concept-possession. Yet in the case of ethical concepts such as *ought*, it seems that one's possession of the concept *ought* could rest on some *prima facie* irrational

disposition of practical reasoning. For instance, a practical reasoner who reasons in accord with the rule Mother Says, a rule telling one to do what one judges one's mother thinks one ought to do, thereby has the concept *ought* and employs it in one's reasoning. But Mother Says is *prima facie* bad as a rule of practical reasoning. Moreover, it seems that someone could also reason in accord with the diametrically opposed rule Mother Forbids, a rule telling one to do what one judges one's mother thinks one ought to avoid.[209] If the standards of practical reason form a coherent set, then it does not seem that each of these rules could be "rational": they could not be authoritative at the same time, for the same person. Yet each of them seems to be a rule such that, in being disposed to follow it, one can possess the concept *ought*.

Moreover, it seems that one can possess *ought* through familiarity with practices of using *ought* that are not cases of practical reasoning at all; through the practices of giving and receiving advice and commands, for example. It certainly seems that children learn the concept *ought* through receiving advice and commands. And if they eventually take heed of such advice and commands in their practical reasoning in some way, as their parents may hope they will, then the disposition first instilled by grasp of the concept *ought* is surely more likely to be something of the Mother Says (or Mother Forbids!) variety than NJI-R. Someone might object that such reasoning is infantile, and if so, then the putative concept *ought* used in such reasoning must be an infantile version of the real concept as well. But it is unclear what reason there is to think that the concept itself is infantile except for the conviction that there must be some good disposition of reasoning that is constitutive of possessing the genuine concept *ought*—a conviction of the very sort at issue in the argument for (b).

In objection, Wedgwood might appeal to a general constitutive ideal of rationality: perhaps one's dispositions must in general approximate rationality in order for one to be intelligible as a thinker at all.[210] But even if this general ideal were true, it would not follow that one cannot possess a given ethical concept in virtue of dispositions that are irrational. One might be put in place as a thinker by one's dispositions with respect to ordinary descriptive concepts and their application. The distinctive trouble

[209] These claims rely in part on the truth of the picture of agency I defend in chapters 2-3.
[210] Wedgwood e.g. 2007: 27, 164; cf. Davidson e.g. 1973b, 1974.

with *ought* and other thin concepts is that they are, precisely, *disputed* concepts: beyond their normative dimension and their thin and trivial descriptive connections to other thin concepts, the conditions of their correct application are not at all clear. While even minimal competence with sortal concepts such as *cat* or *bicycle* plausibly guarantees a degree of correct application—one could not count as competent with the concept *cat* if one went around applying it alternately to mice and to windows—the same does not seem to be true with thin concepts. There precisely seems to be no settled descriptive dimension to *ought* and other thin concepts, beyond the analytical platitudes about their interrelations, to definitively guide one's judgment or reasoning with these concepts. To be sure, there are things that most people might think are (say) irrational, such as acting against one's own interests. But it has not in general seemed to philosophers that this sort of thing can settle the question of correct application in these cases.[211] This is indeed why it is an interesting question what the standards of practical reason really are, in a way that it is not an interesting question what cats or bicycles are.

Nothing in any of this speaks against the more plausible point that *ought* must be a concept that a perfectly rational thinker *could* possess. The point is rather just that there may be several different ways of reasoning with *ought* that are compatible with possession of the concept, and that are not ways of reasoning, even approximately, that the perfectly rational thinker has. Nor need anything in this speak against the intuition that possession of a concept is a "cognitive *power* or *ability*," in and of itself. For as we already saw, in discussing the Entitlement Argument in §4.2, we can account for that intuition simply in terms of the possibilities for knowledge that possessing each concept opens up (pp.159-163). Possessing the concept *ought* potentially puts one in a position to gain new knowledge, but it also puts one in a position to go wrong in new ways. The difference between *ought* and *cat* is that, while there is plenty of room for misapplication of 'ought' in practical reasoning, there is not quite as much room for misapplication of 'cat' in one's theoretical reasoning, once one possesses the concept *cat*. (Though it does

[211] That is, it does not seem true that there are (metaphysically) analytic platitudes about the content of the standards of practical reason, in a way analogous to the way that someone might think there are (metaphysically) analytic platitudes about the content of morality. For the thought that there are such platitudes, and that they can help us discover the metaphysics of morality, see Jackson 1998. For criticism, see e.g. Zangwill 2000.

seem that there is plenty of room for misapplication of 'cat' in one's practical reasoning: if asked what constraints on one's practical reasoning the concept *cat* introduces, I would be mystified as to what to say.)

I have been arguing against Wedgwood's case for (b). There is a corresponding problem with (a) and Wedgwood's argument for it. It is tempting to read the conclusion that "what determines which concept one possesses is just one's dispositions of using the concept in one's reasoning" as involving the claim that different dispositions of reasoning entail different concepts. Certainly this is how we must read the conclusion if it is to rule out cases such as having *ought* while reasoning either just with Mother Says or with Mother Forbids. But Wedgwood's argument only supports the claim that identical roles in reasoning for C and C* mean that C and C* are the same concept. The argument does not support the further claim that possessing a concept C is incompatible with employing it in different types of reasoning from individual to individual. So the argument leaves it open that two individuals' dispositions of practical reasoning with respect to some thin concept could be opposed in exactly the way that rules out the possibility that they could both be rational dispositions.

To be sure, when confronted with cases such as Mother Says and Mother Forbids, we might further deny that either reasoner's grasp of the concept *ought* really *does* depend on their irrational practical dispositions. But since one can have *ought* while having only Mother Says or Mother Forbids as one's practical disposition, this would amount to denying that one's grasp of the concept *ought* must depend upon one's practical dispositions at all. This is, I think, the right response. But it cannot help us argue from the conditions of possessing *ought* to the truth of an ethical view—a view about which practical dispositions lead one to act and reason as one ought.

It seems that similar points apply to other concepts and their involvement in dispositions of practical reasoning—though I cannot of course enumerate all possible cases here. Finally, however, it is worth noting that the argument of §4.3 does in fact make it difficult to sustain claims about the conditions of concept-possession of the sort that the metasemanticist would need. For even if we were ordinarily disposed to reason in accord with, say, NJI-R, the fact that one can coherently doubt NJI-R is still relevant

to showing that the disposition is not a condition of possessing the concept *ought*. For as we saw, one's doubts about the authority of NJI-R might lead one to lose the disposition: since one's doubts are based on a recognition of one's fallibility in making ought-judgments, they might lead one to distrust one's ought-judgments as guides to what one should really do. But in losing the disposition to follow NJI-R, surely the skeptic does not lose the concept *ought*. So having the disposition cannot be a condition of possessing the concept.[212] Given the generality of our previous argument—that one can likewise doubt the authority of any practical rule—it seems that no putative practical norm is safe from this type of extension of the argument. If this is right, then our conclusion stands. Metasemantic and metapragmatic arguments cannot help us vindicate an ethical view.

4.4 CONCLUSION: THE AUTONOMY OF ETHICS

In this chapter, I have argued that neither the conditions of concept-possession, nor those of concept use, can yield a normatively non-question-begging vindication of an ethical view. As far as the conditions of concept-possession and concept use go, it is always coherent for a skeptic to doubt the authority of any putative practical norm.

I said at the outset (§4.1.3) that metasemantic and metapragmatic strategies are one way of attempting to avoid the autonomy of ethics. According to the autonomy of ethics, the only available sound arguments for ethical views are themselves ethically partisan: they inescapably rely on true claims about the content of ethics in supporting further such claims—and in particular, they rely on true claims about the content of ethics that a skeptic can coherently doubt. The partisanship of such arguments is disturbing, since such arguments could not be seen to be sound by anyone who does not already happen to share the right ethical view to at least some extent. Such arguments would leave us with no intellectual

[212] Compare Williamson's remarks on how reflection can lead one to doubt the general validity of MPP, without this entailing that one loses the concept *conditional*, at 2003: 251ff.

response to skeptics about an ethical view, besides futilely repeating the arguments the skeptic rejects, or else engaging in some rousing rhetoric. Even if sound, arguments that are partisan in this way cannot shore up the confidence of anyone, including ourselves, who is genuinely wondering about the credentials of a given ethical view. What is worrying about the autonomy of ethics, then, is not just an idle worry about the possibility of skepticism: it is a worry about the nature of our own intellectual reach to the credentials of any ethical view.

Metasemantic and metapragmatic arguments attempt to avoid the autonomy of ethics by showing that facts about the conditions of concept-possession and concept use entail the truth of some ethical view without begging any normative questions against a skeptic in the process. If my arguments in this chapter are correct, then this way of attempting to avoid the autonomy of ethics cannot succeed. What other ways might there be to avoid it?

One way is constitutivism: the attempt to vindicate ethical views by appeal to the conditions of agency. I have argued against constitutivism in chapters 2-3. The conditions of agency do not implicitly commit us to the authority of any putative practical norm. But might one instead appeal to the conditions of being a theoretical thinker?

It seems that there are only two ways in which the authority of a *practical* rule R, a rule for practical reasoning, could follow from the conditions of being a theoretical thinker. The first is if one must be an agent in order to be a theoretical thinker, and the conditions of agency in turn entail the authority of R. However, even if it is true that one must be an agent in order to be a theoretical thinker, my arguments against constitutivism block this first strategy. It seems that the only other way in which the authority of some practical rule R could be implicit in the conditions of being a theoretical thinker is if each theoretical thinker necessarily has some concept C, or necessarily has some belief B, that somehow commits one to the authority of R. In arguing against metasemantic strategies, I argued that there are no concepts whose possession conditions do commit us to the authority of a putative practical norm. So even if some concepts were necessary to being a theoretical thinker, this would not help us vindicate the ethical

view that R is authoritative. But might there be beliefs B that each theoretical thinker necessarily has that in turn commit us to the authority of R?

It is hard to see how this could be so unless B is a belief that is somehow *about* the content of ethics. If B were a belief about, say, arithmetic, then it is hard to see how its necessity for theoretical thinkers could entail views about good practical reasoning and reasons. But if B is about the content of ethics, then it must be a belief employing normative concepts C, in particular, the thin concepts we investigated in §4.3.2. For as we saw, any ethical view is stated by using those concepts. But then in arguing that the conditions of using thin concepts do not commit us to any specific ethical views, since there are always distancing uses of those concepts in terms of which we can state our skepticism about any given putative practical norm, have we not already argued against the idea that there is some such necessary belief B?

One might think that we have not. For one might suggest that the belief B is not part of the conditions of using the normative concepts C that figure in it; it is just one use. If this were so, then in arguing against the claim that the conditions of using C commit us to the authority of R, we have not yet argued against the claim that B in particular might commit us to R. However, this suggestion is not plausible. If B is a condition of being a theoretical thinker, then B is a use of C that is necessary for using or possessing C at all, since one could not use or possess C unless one were a theoretical thinker; but then B would also be a condition of the use of C in general.

What looks like a more troubling suggestion is that, since we have not explicitly argued against the claim that there is some such belief B, our argument that the conditions of using normative concepts are ethically non-committal is itself incomplete. However, in explaining how it is coherent for skeptics about any given practical norm R to employ thin concepts in explicit denials or suspensions of judgment regarding the authority of R, surely we have thereby explained how it *is* coherent for theoretical thinkers to lack such beliefs B. It is hard to see how one could refute a claim about the necessity of some such belief B for theoretical thinkers *except* by attempting to make sense of how one could lack it. Likewise, it is hard to see how one might positively support the necessity of some such belief B except by attempting

to show that it is not coherent to doubt its content. But then in showing that it is coherent to doubt any claims about the authority of some R, we have shown that it is coherent for theoretical thinkers to lack such a belief B. Of course, it remains possible to assert that there is, too, some such necessary belief B, and that this means that our explanation of how one could lack it must be somehow faulty. But mere assertion does not fare well against explanations of the coherence of skepticism.

If this is right, then in arguing against metasemantic and metapragmatic strategies, we have simultaneously argued against the claim that something in the conditions of being a theoretical thinker might entail the authority of R. So one can be an agent, a theoretical thinker, and a possessor and a user of thick and thin ethical terms, without thereby committing oneself to the authority of any putative practical norm R—and in fact, while actively doubting the authority of any putative practical norm R. Therefore, the authority of any putative practical norm R can be coherently doubted.

The autonomy of ethics follows. Since there is nothing that implicitly commits a skeptic about R to R's authority, any argument for the authority of R must beg some questions against a possible skeptic about R. Ethical argumentation must proceed in inescapably partisan terms.

Of course, some questions begged against the skeptic might be about putative grounding assumptions for practical norms: for instance, someone might claim that the standards of practical norms are just the standards of what it is to be good *qua* human being, and on this general basis try to derive more specific conclusions about what the standards of practical reason are.[213] Nonetheless, such grounding assumptions must themselves be ethically committal: they consist in principles to the effect that the standards of practical reason just are the standards of goodness *qua* human being. (And not, say, the standards of goodness *qua* thief.) They are general ethical principles, and so principles that a skeptic could coherently doubt. So even though some of the ethical claims in terms of which ethical argumentation may proceed are very general, our ethical argumentation is no less partisan for that.

---

[213] For this type of strategy, see Foot 2001.

How much should this conclusion disturb us? It may seem to make our views about how to live rationally arbitrary, and any declaration of allegiance to an ethical view a case of mere dogmatism. But of course, inescapable partisanship need not amount to dogmatism in any sense entailing resistance to reflective change in view. The serious philosophical questions here concern whether and in what sense reflection internal to ethics can be shown to suffice for the possibility of ethical knowledge; and whether and how we can make sense of ethics as objective without being able to fully ground its putative truths in anything. In particular, if we wish to maintain that there are objective standards of practical reason, it seems that we need some explanation of what objectivity in the area consists in that distinguishes our view from the views of various types of non-factualist about practical norms. This latter task looks metaphysical. And the difficulty that my argument poses for the task is that it can be hard to see how there could be a metaphysical explanation of the objectivity of practical norms that did not at the same time amount to a metaphysical *criterion* by which we could tell which putative practical norms are the objectively authoritative ones; and my argument has been, in effect, that no such purely metaphysical criterion is available.

BIBLIOGRAPHY

Allison, H. 1990. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Anscombe, G. E. M. 1957. *Intention*. Cambridge, MA: Harvard University Press.

Anscombe, G. E. M. 1995. Practical Inference. Reprinted in *Virtues and Reasons,* eds. Hursthouse, Lawrence & Quinn. Oxford: Clarendon Press, 1995.

Aquinas, *Summa Theologica*. Transl. Fathers of the Dominican Province. Christian Classics, 1981.

Aristotle. *Nicomachean Ethics* (NE). Transl. & Intr. S. Broadie & C. Rowe. Oxford: Oxford University Press, 2002.

Arpaly, N. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.

Boghossian, P. 2001. How Are Objective Epistemic Reasons Possible? In Boghossian, P. *Content & Justification: Philosophical Papers*. Oxford: Oxford University Press, 2008.

Boghossian, P. 2003a. Blind Reasoning. Reprinted in Boghossian, P. *Content and Justification: Philosophical Papers*. Oxford: Clarendon Press, 2008.

Boghossian, P. 2003b. Epistemic Analyticity: A Defense. Reprinted in Boghossian, P. *Content and Justification: Philosophical Papers*. Oxford: Clarendon Press, 2008.

Boghossian, P. 2008a. Epistemic Rules. In Boghossian, P. *Content and Justification: Philosophical Papers*. Oxford: Clarendon Press, 2008.

Boghossian, P. 2008b. *Content and Justification: Philosophical Papers*. Oxford: Clarendon Press, 2008.

Bratman, M. 2003. A Desire of One's Own. Reprinted in Bratman, M. *Structures of Agency: Essays*. Oxford: Oxford University Press, 2007.

Burge, T. 1998. Reason and the First Person. In *Knowing Our Own Minds*, eds. C. Wright, B. Smith, & C. McDonald. Oxford: Clarendon Press, 1998.

Cappelen, H. & LePore, E. 2005/2009. Quotation. In *Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta.

Carroll, L. 1995. What The Tortoise Said to Achilles. *Mind*, New Series 104: 691-693. Orig. 1895, *Mind* 4: 278-80.

Davidson, D. 1973a. Freedom to Act. Reprinted in Davidson, D. *Essays on Actions and Events*. Oxford: Oxford University Press, 1980.

Davidson, D. 1973b. Radical Interpretation. In Davidson, D. *Inquiries into Truth and Interpretation.* Oxford: Clarendon Press, 1984.

Davidson, D. 1974. Belief and the Basis of Meaning. In Davidson, D. *Inquiries into Truth and Interpretation.* Oxford: Clarendon Press, 1984.

Dreier, J. 1997. Humean Doubts about the Practical Justification of Morality. In *Ethics and Practical Reason*, eds. Cullity & Gaut. Oxford: Clarendon Press, 1997.

Dreier, J. 2001. Humean Doubts about Categorical Imperatives. In *Varieties of Practical Reasoning*, ed. E. Millgram. Cambridge, MA: MIT Press, 2001.

Ekstrom, L.W. 2005. Autonomy and Personal Integration. In *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, ed. J.S. Taylor. Cambridge: Cambridge University Press, 2005.

Finlay, S. 2006. The Reasons That Matter. *Australasian Journal of Philosophy* 84: 1-20.

Finlay, S. 2007. Responding to Normativity. In *Oxford Studies in Metaethics*, Vol. 2, ed. R. Shafer-Landau. Oxford: Oxford University Press, 2007.

Foot, P. 1972. Morality as a System of Hypothetical Imperatives. Reprinted in Foot, P. *Virtues and Vices.* Oxford: Oxford University Press, 2002.

Foot, P. 1994. Rationality and Virtue. Reprinted in Foot, P. *Moral Dilemmas*. Oxford: Oxford University Press, 2002.

Foot, P. 2001. *Natural Goodness*. Oxford: Clarendon Press.

Frankfurt, H. 1975. Three Concepts of Free Action. Reprinted in Frankfurt, H. *The Importance of What We Care About*. Cambridge: Cambridge University Press, 1988.

Frankfurt, H. 1977. Identification and Externality. Reprinted in Frankfurt, H. *The Importance of What We Care About*. Cambridge: Cambridge University Press, 1988.

Frankfurt, H. 1987. Identification and Wholeheartedness. Reprinted in Frankfurt, H. *The Importance of What We Care About*. Cambridge: Cambridge University Press, 1988.

Harman, G. 1999. *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press.

Hooker, B. 1987. Williams' Argument against External Reasons. *Analysis* 47: 42-44.

Hume, D. *A Treatise of Human Nature.* D.F. Norton & M. J. Norton, eds. Oxford: Oxford University Press, 2000.

Hursthouse, R. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.

Jackson, F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.

Jollimore, T. 2005. Why Is Instrumental Rationality Rational? *Canadian Journal of Philosophy* 35: 289-307.

Kant, I. *Groundwork of the Metaphysics of Morals.* Transl. & ed. H.J. Paton. London: Routledge, 1948.

Kant, I. *Religion Within the Boundaries of Mere Reason.* In A.W. Wood & G. Di Giovanni, transl. & eds. *Religion and Rational Theology*: *The Cambridge Edition of the Works of Immanuel Kant.* Cambridge: Cambridge University Press, 1996.

Kavka, G. S. 1983. The Toxin Puzzle. *Analysis* 43: 33-36.

Kolodny, N. 2008. Why Be Disposed to Be Coherent? *Ethics* 118: 437-463.

Korsgaard, C. M. 1986a. Aristotle on Function and Virtue. *History of Philosophy Quarterly* 3/3: 259-79.

Korsgaard, C. M. 1986b. Skepticism about Practical Reason. *The Journal of Philosophy* 83: 5-25.

Korsgaard, C. M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, C. M. 1997. The Normativity of Instrumental Reason. In Cullity & Gaut, eds. *Ethics and Practical Reason*. Oxford: Clarendon Press, 1997.

Korsgaard, C. M. 2008a. Aristotle's Function Argument. In Korsgaard, C. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology.* Oxford: Oxford University Press, 2008.

Korsgaard, C. M. 2008b. Acting for a Reason. Reprinted in Korsgaard, C. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology.* Oxford: Oxford University Press, 2008.

Korsgaard, C. M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

McDowell, J. 1979. Virtue and Reason. Reprinted in McDowell, J. *Mind, Value, & Reality*. Cambridge, MA: Harvard University Press, 1998.

McDowell, J. 1981. Non-Cognitivism and Rule-Following. Reprinted in McDowell, J. *Mind, Value, and Reality.* Cambridge, MA: Harvard University Press, 1998.

McDowell, J. 1995a. Might There Be External Reasons? In *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, eds. Altham & Harrison. Cambridge: Cambridge University Press, 1995.

McDowell, J. 1995b. Eudaimonism and Realism in Aristotle's Ethics. In R. Heinaman, ed. *Aristotle and Moral Realism*. Boulder: West View Press, 1995.

McDowell, J. 1996. Deliberation and Moral Development in Aristotle's Ethics. In S. Engstrom & J. Whiting, eds. *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*. Cambridge: Cambridge University Press, 1996.

Nagel, T. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Peacocke, C. 1998. Implicit Conceptions, Understanding and Rationality. *Philosophical Issues* 9: 43-88.

Pettit, P. & Smith, M. 1990. Backgrounding Desire. *The Philosophical Review* 99: 565-592.

Pettit, P. & Smith, M. 1996. Freedom in Belief and Desire. *Journal of Philosophy* 93: 429-449.

Plato, *Republic*. Transl. Grube, G.M.A, Revised by Reeve, C.D.C.; Hackett Publishing Company, Inc. 1992.

Prior, A. N. 1960. The Runabout Inference-Ticket. *Analysis* 21: 38-39.

Quinn, W. 1993. Putting Rationality in Its Place. In W. Quinn, *Morality and Action*. Cambridge: Cambridge University Press, 1993.

Raz, J. 2005a. The Myth of Instrumental Rationality. *Journal of Ethics & Social Philosophy* 1: 2-28.

Raz, J. 2005b. Instrumental Rationality: A Reprise. *Journal of Ethics & Social Philosophy*, Symposium I, Dec. 2005.

Ridge, M. & Barandalla, A. 2010. Critical Notice: Function and Self-Constitution: How to make Something of Yourself without Being All That You Can Be. A Commentary on Christine Korsgaard's *The Constitution of Agency* and *Self-Constitution*. *Analysis* Advance Access, July 29 2010.

Rödl, S. 2007. *Self-Consciousness*. Cambridge, MA: Harvard University Press.

Rödl, S. 2010. The Form of the Will. In *Desire, Practical Reason, and the Good*, ed. Tenenbaum, S. Oxford: Oxford University Press, 2010.

Schaffer, J. 2009. On What Grounds What. In *Metametaphysics: New Essays on the Foundations of Ontology*, eds. Manley, Chalmers & Wasserman. Oxford: Oxford University Press, 2009.

Schroeder, M. 2007a. The Humean Theory of Reasons. In *Oxford Studies in Metaethics*, Vol. 2, ed. R. Shafer-Landau. Oxford: Oxford University Press, 2007.

Schroeder, M. 2007b. *Slaves of the Passions*. Oxford: Oxford University Press.

Schroeter, F. 2005. Normative Concepts and Motivation. *Philosophers' Imprint* 5, Vol.3: 1-23.

Schueler, G. F. 2009. The Humean Theory of Motivation Rejected. *Philosophy and Phenomenological Research* 78: 103-122.

Setiya, K. 2007. *Reasons without Rationalism*. Princeton: Princeton University Press.

Setiya, K. 2010. Sympathy for the Devil. In *Desire, Practical Reason, and the Good,* ed. S. Tenenbaum. Oxford: Oxford University Press, 2010.

Smith, M. 1987. The Humean Theory of Motivation. *Mind*, New Series, 96, No. 381: 36-61.

Smith, M. 1987-8. Reason and Desire. *Proceedings of the Aristotelian Society* 88: 243-258.

Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.

Smith, M. 2009. The Explanatory Role of Being Rational. In *Reasons for Action,* eds. Sobel & Wall. Cambridge: Cambridge University Press, 2009.

Smith, M. 2010. Beyond The Error Theory. In *A World Without Values: Essays on John Mackie's Moral Error Theory*, eds. R. Joyce & S. Kirchin. Springer, 2010.

Stocker, M. 1981. Values and Purposes: the Limits of Teleology and the Ends of Friendship. *The Journal of Philosophy* 78: 747-765.

Svavarsdóttir, S. 1999. Moral Cognitivism and Motivation. *Philosophical Review* 108: 161-219.

Velleman, J.D. 1992. What Happens When Someone Acts? Reprinted in Velleman, J.D. *The Possibility of Practical Reason.* Oxford: Oxford University Press, 2000.

Velleman, J.D. 1996. The Possibility of Practical Reason. Reprinted in Velleman, J.D. *The Possibility of Practical Reason.* Oxford: Oxford University Press, 2000.

Velleman, J.D. 1997. Deciding How to Decide. Reprinted in Velleman, J.D. *The Possibility of Practical Reason.* Oxford: Oxford University Press, 2000.

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.

Vogler, C. 2002. *Reasonably Vicious*. Cambridge, MA: Harvard University Press.

Wallace, R. J. 1990. How to Argue about Practical Reason. *Mind* 99: 355-85.

Watson, G. 1975. Free Agency. Reprinted in Watson, G. *Agency and Answerability: Selected Essays*. Oxford: Oxford University Press, 2004.

Watson, G. 1987. Free Action and Free Will. Reprinted in Watson, G. *Agency and Answerability: Selected Essays*. Oxford: Oxford University Press, 2004.

Wedgwood, R. 2002. Practical Reason and Desire. *Australasian Journal of Philosophy* 80: 345-358.

Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.

Williams, B. 1979. Internal and External Reasons. Reprinted in Williams, B. *Moral Luck*. Cambridge: Cambridge University Press, 1981.

Williams, B. 1985. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.

Williamson, T. 2003. Understanding and Inference. *Proceedings of the Aristotelian Society, Supplementary Volume* 77: 249-93.

Wittgenstein, L. 1953. *Philosophical Investigations*. Blackwell, 3rd edition, 2003.

Zangwill, N. 2000. Against Analytic Moral Functionalism. *Ratio (new series)* 13: 275-286.