



Acting From Thought About Action

Citation

Fix, Jeremy D. 2016. Acting From Thought About Action. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:33840677>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

© 2016 Jeremy David Fix

All rights reserved

Acting from Thought about Action

Abstract

Human action is unique. It is metaphysically unique because we can act self-consciously. It is normatively unique because we are subject to prudential, moral, and rational standards in action, whereas other agents are not. What is the relationship between these aspects of our action? I defend a constitutivist view about the nature of normativity whose basis is the view that a capacity is such that a single principle describes its nature and is thereby normative for its development and exercises. I argue that this view answers basic questions in metaethics and action theory and allows us to understand distinct notions of possibility essential to various conditions on our agency and theses about our agency. I then turn to questions about practical reason in particular, arguing against intellectualist and instrumentalist views that in different ways try to explain practical reason in terms of theoretical reason. I finally turn to develop a positive, non-reductive view of practical reason as the self-conscious kind of will. I explain how the general constitutivism of capacities affects our account of the self-conscious will in general, our determinate version of it, and the relationship between them.

Acknowledgments

At various points while writing this dissertation, I did not expect to survive my PhD, let alone complete it. I owe much of the credit for this work to the four members of my committee and to various members of the faculty of the Harvard philosophy department, past and present. I cannot say everything that I wish to say about everyone I wish to acknowledge in this space. Let me just say that my sense is that Harvard is one of the only departments in which I could have pursued these interests in these ways, not only because of the people here sympathetic with some of my views but more importantly because of the people who are not. Although many faculty members, including at least one of my committee members, probably think that the substance of my views are not just wrong but headed in entirely the wrong direction, everyone has always been supportive and in fact positively encouraging about my pursuing them and about my development of them. This kind of openness, not just to views that you think are likely to be true but to views that you think are almost certainly false, is unique in my experience. I cherish it.

Chris Korsgaard is a perpetual inspiration, as deep and searching in her advising as she is in her own work, as charitable with her time and thoughts as she is when interpreting the great figures of the past so central to her work and the view of philosophy she passes onto her students. Ned Hall no doubt did not expect to advise a dissertation about the nature of practical reason, let alone to bear witness to the phoenix-like rebirth of an Aristotelian view of the living in terms of a metaphysics of the capacities of the living and their place in an explanation of the nature of normativity. Dick Moran's work on self-knowledge drew me to graduate study at Harvard. His support for the content of my work combined with criticism and advice on its early presentations helped me see through the confusion, my own and others, engendered by my first attempts to express these ideas, separating the problems from the merely problematic formulations. Matt Boyle and Doug Lavin are forever linked in my mind. They more than anyone else have influenced my thought on these issues, my approach to philosophical questions generally, how I interact with other philosophers and other views, and everything else that is part of my philosophical disposition. From a graduate seminar in my first-semester to discussions at conferences during this last summer, they have shaped my graduate experience and my sense of how a graduate

department should function. Selim Berker is the ideal reader, charitable and critical in measures equaled only by his open-mindedness and excitement about any honest and searching attempt to address and answer a philosophical question.

If one question characterizes the approach to philosophical issues of all of my advisors, formal and informal alike, it is about how we—the ‘we’ of philosophy, of humanity—should approach these issues and the way that often tacit answers to that question shape subsequent views of the subject matter of interest. The Practical Philosophy Graduate Student Workshop was often the site of my first approach to the issues that I discuss in this work and to issues that I learned on those occasions I am better off leaving for later if not abandoning entirely. I thank the members of that workshop throughout the many years for helping me with my work and, more importantly, for sharing their work with me. I might have been able to finish this dissertation without the deadlines that presentations provided. I might have been able to finish it without having the privilege to read and discuss everyone else’s work in order to pleasantly distract me from the difficulties of my own. I and it would, though, have been all the worse for it.

I presented versions of “Intellectual Isolation” at the Safra Center Graduate Fellows Seminar and the 4th Annual TVA Conference at the University of Tennessee, Knoxville. I presented versions of “The House of Goodness” at the Safra Center Graduate Fellows Seminar, an Nth Year Seminar at Harvard, the Theoretical Philosophy Graduate Student Workshop, and at the University of California, Riverside. Thanks much to the audiences at these talks.

Thanks much to the graduate students who have spent time at Harvard and spent much time with me discussing philosophy and much else. Rachel Achs, James Bondarchuk, Colin Chamberlain, Byron Davies, Johann Frick, Micha Glaeser, Chandler Hatch, Michael Kessler, Allison Kuklok, Enoch Lambert, David Langlois, Celine Leboeuf, Barry Maguire, Paul Marcucilli, Emily McWilliams, Elizabeth Miller, Oded Na’aman, Kranti Saran, Paul Schofield, Jiewuh Song, Aleksy Tarasenko-Struc, and Kate Vrendenburgh. You have been much better friends and interlocutors than my conduct always warranted.

Thanks finally and most importantly to Thomas Pendlebury, who for the last two years has spent more time helping me become a better philosopher and a better person than I could reasonably expect

from a partner. I would spend a bunch of words extolling his virtues, embarrassing myself given that my literary talents are far too minimal to openly express such emotions in a clear and moving way. Fortunately, he strongly dislikes contemporary norms for acknowledgments. The best way for me to acknowledge him, then, is to leave off at this point.

Table of Contents

	Introduction	1
1.	The House of Goodness	5
2.	Practical Possibility	33
3.	Intellectual Isolation	97
4.	The Empty 'I'	138
5.	I do. What Happens?	167
6.	The Substance of Constitutivism	195
	Bibliography	254

INTRODUCTION

HUMAN BEINGS are reflective animals. Philosophy is one way that we exercise this capacity. Yet reflection is not limited to abstract tasks. Reflection is an essential part of our lives, with an essential role in our exercises of theoretical reason in which we come to know the world and in our exercises of practical reason which we come to change the world.

This dissertation is an essay in practical philosophy, an exercise of reflective reason about the nature of practical reason. As far as we know, we are the only animal with this capacity to act self-consciously. As far as we know, we are the only animal subject to moral and prudential requirements in our action as such. I shall argue that the metaphysical uniqueness of this capacity explains its normative uniqueness. We are subject to moral and prudential requirements because of the nature of our capacity to act.

I first explain and defend a metaphysics of capacities of the living that says that a single principle describes the nature of a capacity and is thereby normative for its development and exercises. This metaphysics establishes a version of *constitutivism*, the view that what something is determines what it ought to be, at least with respect to things subject to internal normative standards. I argue that it meets a basic challenge of action theory about how to account for our action within a general metaphysics of reality without reducing our action to something else. I argue that it also meets a basic challenge of metaethics about how to account for scope and nature of normativity and the different kinds of normative standards that apply to different kinds of things.

This approach places practical reason within a broader metaphysics of capacities. It does not by

itself offer an account of practical reason. With only minimal assumptions about the nature of practical reason that are common ground for most practical philosophers, though, this metaphysics is able to explain the notions of possibility within various conditions of agency and theses about our agency. *Ought implies Can*, *The Error Condition*, and *The Principle of Alternate Possibilities* link what I ought to do, whether I am subject to a normative standard, and whether I am responsible for acting in some way with what *I can do*. Likewise, although the thesis of the guise of the good and the claim that to act intentionally is to act knowingly do not feature modal terminology, objections to them turn on counterexamples that show that I *can* act in some way even if I do not think that so acting is good and even if I do not know what I am doing or why I am doing it. I argue that the metaphysics explains how the notions of possibility in the conditions of agency differ from each other in terms of two distinctions that follow from it. On the one hand, the metaphysics distinguishes possessing a capacity because I am a human being from my state of development of that capacity. On the other hand, the metaphysics divides what is possible with respect to a capacity into two sub-classes, the perfections of an exercise of a capacity that constitute such exercises as good and the imperfections that constitute such exercises as bad. Different conditions use different combinations of these distinctions in the notion of possibility in them. I also argue that the guise of the good and other claims about our agency are claims about the nature of our capacity to act. They are thereby normative for exercises of that capacity, not universal generalizations about those exercises. The putative counterexamples are then in fact evidence for the theses in question, at least so long as the counterexamples are per se imperfect exercises of the capacity.

While minimal assumptions about the nature of practical reason are enough to offer a view of those conditions and theses within the metaphysics of capacities, this metaphysics does not provide an

account of practical reason. An account of it must instead clarify what practical reason is a capacity to do. I defend a view of practical reason as distinct in kind or form from theoretical reason. Practical reason is our kind of will, our capacity to act.

Before I explain and defend that view, though, I argue against two common views of practical reason and one familiar view of our action that are incompatible with it. I first argue against two views that in different ways try to explain practical reason in terms of theoretical reason. Instrumentalists think that its only roles are to form means-ends beliefs and evaluate ends in light of other ends. A non-rational capacity gives the ends. It works out consequences from them. Intellectualists, in contrast, think that reason supplies ends for our action. These exercises of reason are theoretical, though, distinguished from others by their objects: reasons to act. On both views, then, all exercises of reason are theoretical in nature. Neither view is adequate. Against intellectualism, I argue that if practical reason is a species of theoretical reason, practical judgments cannot have authority over the will. Without that authority, though, I cannot act from my representation of a reason to act. I then argue that if I cannot act from those representations, those reasons and a species of theoretical reason about them cannot exist. Against instrumentalism, I argue that its claims about how to decide are ambiguous between descriptive claims and normative claims. Neither disambiguation does the needed explanatory work.

I then argue against the metaphysical reduction of agency prominent in action theory under the name *the causal theory of action*. However, I argue that this theory cannot meet the explanatory burdens of a metaphysical reduction. Every account of action must invoke a means-end belief, which is a belief about what I can self-consciously do in order to achieve my end. A metaphysical reduction of action cannot invoke the idea of what I can self-consciously do, though, since that is what it means to explain.

Hence, a metaphysical reduction of our action is impossible.

I finally turn to sketching the view that practical reason is our kind of will. I derive an instrumental requirement and a prudential requirement from the idea of acting self-consciously. Most philosophers assume that this kind of view must say that we are subject to moral requirements only if we can also analytically derive them from the concept of self-conscious action. I argue, though, that this kind of derivation is unnecessary and does not fit with the idea that our capacity to act is the capacity of a living being just like any other. Just as the genus of a capacity for digestion can have different species whose bearers are to eat and drink different things and process them in different ways, so the genus of a capacity for self-conscious action can have different species whose bearers are subject to different normative standards in their exercises of those capacities as such. Just as we cannot analytically derive all the substantive normative standards that govern our digestion from the concept of digestion, so we cannot derive all the substantive normative standards that govern our action from the concept of self-conscious action. Practical reason is thus a genus with different species, and beings with different species are subject to different normative standards. I explain why this fact should not worry us when we wonder whether we are moral beings—that is, whether we are subject to moral and prudential requirements in action as such. This explanation shows that knowledge of the principle of our will is reflective knowledge that reflects the developed ethical character of the agent in question.

THE HOUSE OF GOODNESS

[T]he irrational is not merely the non-rational, which lies outside the ambit of the rational; irrationality is a failure within the house of reason. ... Irrationality is a ... rational process or state ... gone wrong. (Donald Davidson [1982] 169)

1. REALITY AND GOODNESS

THE REAL is the Good. So an ancient view says, anyway, but we are far from antiquity. David Hume says that no ‘ought’ follows from an ‘is’. Most philosophers, Hume among them, take a metaphysical division to undergird this logical gap. And who can deny it? Much of reality is rotten, much good unrealized.

Of course, the ancients knew about all that. How, then, can they identify reality and goodness? I will not address this question. I shall, though, develop and defend *constitutivism*, which in its own way links reality and goodness. It says, roughly, that what something is determines what it should be, at least with respect to certain kinds of things. As I develop the view, its basis is a metaphysics of capacities which says that a single principle describes the nature of a capacity and is thereby normative for its exercises.¹ Capacities are thus a type of potentiality that divide their actualizations into the good or bad, perfect and defective, successful and unsuccessful. Why? Because capacities are potentialities to *do* things, to be active in ways that only living things can be active. To be active in this sense is to exercise a capacity. Doing something is thus by nature subject to a standard of goodness. The Real and the Good are united in this way in these things.

I shall argue that this metaphysics lets constitutivism meet what I call *the basic challenges of action theory and metaethics*. Let me say a bit about these challenges. Roughly, the first is to account for our action within a general metaphysics. We are of and act in the same world in which everything else exists

1. I here only develop the aspects of this metaphysics of capacities needed to meet the basic challenges of metaethics and action theory. In particular, I ignore the fact that many of our capacities by nature develop over time. I explain the significance of this aspect of the metaphysics for practical philosophy in “Practical Possibility”.

and happens. An account of our action must not make it seem out of this world. Why might it? Well, think about how action theorists set up the challenge. Brian O'Shaughnessy claims that our

action seems like a leak from another realm or world into this world, a leak or intervention ... such as God would effect were He able to effect change in the world without transgressing the Laws of Nature. ... By action we irreducibly alter the state of the universe This is *creation*. We are ultimate sources of change in the environment. ... If I am drawing a line, it is not in the final analysis the product of a physical force: it is ... but from my body as vehicle from another realm: that of (my) reason and purpose. We stand within and without physical nature. (O'Shaughnessy [1980/2008] 317-8)

Christine Korsgaard similarly says that to

be an agent ... is to be a self-conscious causality: to think to yourself that you will bring about a certain end, and somehow, through that thought, to bring it about. When you think of yourself as an agent, you think that your effects are your own, and had you not taken thought to realize them, they would not, other things being equal, have happened at all. So you are their cause, and these effects are yours, and the world is different, because you've made it so. There's a touch of the divine in being an agent. (Korsgaard [2009] 84)

She later claims that the 'problem of action' is to explain how to "act—that is, *if* such a thing is possible at all—is to insert yourself—your first-personal, deliberating self—into the causal network" (ibid. 86). J.

David Velleman likewise claims that we

don't seem to be adrift in the flow of events: we seem to intervene in it, by producing some events and preventing others. Yet our intervention invariably consists in thoughts and bodily movements, which either happen by chance or are caused to happen by other thoughts and movements, which are themselves events taking place in our minds and bodies. Our intervening in the flow of events is just another part of that flow. So how can it count, after all, as an intervention—or, for that matter, as ours? (Velleman [2000_b] 5)

We intervene in the flow of events of the world, insert ourselves into it, alter its history, leak from one world into another, or stand within and without nature. We are ultimate creators, first-causes or prime movers, tiny deities who partly determine the course of the world through thought realized in action.

These evocative images ring true but also perplex. We act self-consciously, and this type of self-

determination differs from everything else in the world. Yet these images make our action seem like a fantasy. How can we understand it and thus ourselves without it seeming otherworldly—unnatural and mysterious?

Many action theorists try to answer this question by addressing a specific version of the basic challenge. They treat action theory as a bit of applied metaphysics. They posit an ontology for the ‘general metaphysics’ without self-conscious self-determination and try to reduce our action to it. While I criticize this view elsewhere, my aim here is constructive. We can meet this challenge, I shall argue, by understanding how our action transposes a form of explanation proper to all activities of the living into the key of self-consciousness. Instead of falling outside of a metaphysics that accounts for everything else, the metaphysics of our action is a determinate form of the metaphysics of anything the living does. Since a mark of the living is that we are active, our action is as of the world as life itself.

What about the basic challenge of metaethics? Roughly, this challenge is to explain the nature and structure of normativity. Normativity is everywhere. I can act well or badly, my cat can act well or badly, we can digest well or badly, and so on for other activities of the living. Likewise, artifacts can be good or bad tokens of their kind, and just about anything can be good or bad for whatever uses the living make of it. What are the different kinds of normative standards, what explains them, and how do they relate to each other? Of importance for practical philosophy, the standard for our capacity to act can play a role in its exercises that standards for other capacities cannot play in their exercises. Standards apply *to* us, as with everything, but they can also be applied *by* us. We can evaluate stuff according to them. More importantly, we can guide ourselves by them in action, including by the standard that governs our action as such. What explains this difference?

Here again, I shall argue, the key is that the metaphysics of our action is a determinate form of the metaphysics of anything the living does. To be active is to exercise a capacity, and the nature of a capacity establishes a standard for its exercises. We make use of things in exercises of capacities, and the standards that apply to them derive from the nature of the capacities being exercised. In this way, the capacities of the living ground all normativity. Our capacity to act is self-conscious, though, which explains the unique role that its standard can play in its exercises.

I shall approach these issues indirectly. Constitutivism is about the relationship between metaphysics and normativity. Most discussion of it focuses on our agency and which standards, if any, the nature of our agency can establish for our action. Different constitutivists claim to derive different standards from the nature of action. Critics challenge these claims about our agency in particular, either arguing that a specific derivation fails or arguing that no such derivation can work for our action. I think that taking constitutivism seriously requires more general metaphysical reflection, which reveals a more complex, flexible, and interesting view than is often recognized. Its promise and structure come out best in explicit contrast with common assumptions in practical philosophy when we recognize our action as one activity of one kind of living being and explicitly contrast. We must pull back our gaze from the most important matters in practical philosophy in order to see them for what they are.

To that end, I will first articulate a view about the relationship between goodness and our action that lies behind the methodology of much practical philosophy. While I think that this assumption gets in the way of meeting the basic challenges, I shall not argue the case. I instead will show that a common objection to constitutivism just asserts that it is incompatible with this assumption. Exposing this assumption thus lets me formulate constitutivism in a way that foregrounds distinctive aspects of its

explanatory structure. I shall then show how to meet the basic challenges of action theory and metaethics within that framework. I will not, though, offer an account of the principle of our capacity to act. My goal is to talk you into a way of thinking about normativity and the activity of the living that can meet the basic challenges independent of any substantive views about our action that might divide constitutivists. I shall refer to unique aspects of our agency only when and to the extent that it is needed in order to meet those challenges. I leave discussion of the principle of our capacity to act for other work, which I begin but in no way complete in the “The Substance of Constitutivism”.

2. A MOOREAN TRAGEDY

Contemporary practical philosophy is a tragedy. Although goodness and our action belong together, practical philosophers separate them. The separation is wrong, the consequences dire. Like any such tragedy, the story begins with George Edward Moore.

Moore claims that our action is “the commonest and most generally interesting object of ethical judgments”, but he thinks that goodness and our action have no essential relationship to each other (Moore [1903] §2). For one thing, “‘good conduct’ is a complex notion: all conduct is not good, for some is certainly bad and some may be indifferent” (ibid.). Since some actions are not good, he concludes that goodness is not part of the nature of our action. Good and bad actions exemplify the kind. Action theorists thereby should not worry about goodness. For another thing, “other things, beside conduct, may be good; and if they are so, then, ‘good’ denotes some property, that is common to them and conduct” (ibid.). Since some good things are not actions, Moore concludes that the nature of goodness is independent of the nature of our action even when it is one of our actions that is good.

Tokens of different kinds have the same property of goodness. Metaethicists thereby should not worry about our action.

This simple and seductive argument turns on two metaphysical assumptions, one about kinds and the other about goodness. Take the move from the badness of some actions to the claim that goodness is not part of the nature of our action. This move is sound only if a property is not part of the nature of a kind if it does not inhere in every token. Put otherwise around, it is sound only if only *necessary properties* of a kind are part of its nature.² Call this view about kinds *necessitarianism*.

Now take the move from the diversity of things that can be good to the claim that the nature of goodness is independent of the nature of our action even when it is one of our actions that is good. This move is sound only if the same property of goodness inheres in the same way in tokens of different kinds regardless of the nature of the kinds in question. To understand this idea, think about how the property of redness—or, if you must, a shade of red—differs from the property of having an organ. The redness of something does not depend on the nature of the thing in question, which is why you need not know about wires in order to know what it is for one to be red. In contrast, how something has an organ depends on its nature, which is why you need to know about cats and cows in order to know what it is for one to have an organ. After all, human brains, cat hearts, cow stomachs, and whale lungs differ substantially while being ways of having an organ. The ‘organ-ness’, as it were, of these things differs in a way that the crimson of a blotch and a piece of fabric does not. Moore assumes that goodness is like

2. This claim is compatible with the view that some necessary properties of a kind are not part of its nature. Much current debate about the nature of essence assumes that *only* necessary properties of the kind are part of its nature and focuses on whether all or only some of the necessary properties of the kind are part of its nature. This debate is offstage for my purposes because constitutivism says that some contingent properties of kinds are part of the nature of some kinds. Hence, the assumption is wrong with respect to kinds that embody the constitutivist metaphysics.

redness. Call this view about goodness *separatism*.

I refer to the conjunction of necessitarianism and separatism as *divisionism* since both divide the nature of goodness from the nature of kinds whose tokens can be good, at least assuming that any kind that can have good tokens can also have bad ones. Most practical philosophers are divisionists. At least, their methodology makes sense in light of it. The case is clearer with action theorists, so start with them.

Action theorists tend to think that goodness does not have a role in an account of our action. Most try to account for our action in terms of a set of mental states efficiently causing movements. Many others instead say that the agent herself efficiently causes the movements. On each view, the notion of causation is generic, the kind you find in tornadoes and continental drift. The nature of willing, whether reduced or not, can be fully exemplified regardless of whether the willed action is good or bad. For most action theorists, then, goodness has no more role in an account of our action than it has in an account of hurricanes or hail storms. 'Good human action' is then a 'complex notion' that combines independent parts.

Metaethicists, for their part, tend to go for an independent handle on "what is good in general; hoping, that if we can arrive at any certainty about this, it will be much easier to settle the question of good conduct: for we all know pretty well what 'conduct' is" (ibid. §2). They might not do it in these terms. Take the view that the basic normative element is a consideration counting in favor of something for some being. *That* is the general account of a reason. Reasons theorists do not first account for reasons to act, reasons to believe, and the rest and then abstract in order to get a general account. They instead start with the general account and try to explain what counts in favor of what for whom. Now take the view that something is good just if favorably responding to it in some way is fitting. *That* is the general

account of goodness. Fittingness theorists do not first account for good human action, good cat digestion, and the rest and then abstract in order to get a general account. They instead start with the general account and try to explain which favorable responses are fitting for which beings to have to which things. These accounts might be naturalist or non-naturalist, expansive or minimalist, reasons-first or goodness-first, buck-passing or buck-stopping, and so on. Regardless of the details, they try to explain good human action in terms of the general account of the normative property or relation and an independent account of our action. 'Good human action' is again a 'complex notion' that combines independent parts.

Most practical philosophers thus do not make much of the fact that our actions can be good. Accounts of goodness and our action must be compatible with it in the same way that accounts of rocks and redness must be compatible with the possibility of red rocks. These accounts, though, are independent of each other. This methodology makes sense only if goodness is not part of the nature of our action and only if the nature of goodness is independent of the nature of our action even when it is one of our actions that is good. The second claim just is separatism about goodness. And although necessitarianism is not the only way for goodness to not be part of the nature of our action, it is simplest explanation for why it is not and embodies a widespread assumption about the nature of kinds. Hence, the methodology of practical philosophy belies a commitment to divisionism.

3. ACTION AND ERROR

To fully defend constitutivism requires rebutting divisionism. My interest in divisionism in this paper is more limited. I think that it sneaks into discussion of constitutivism because much practical philosophy

rests on it. Consider *the problem of bad action* for constitutivism. This problem is an illusion, but the way that some constitutivists talk conjures it. They say that a principle is descriptive of and normative for an activity. This view is superficially paradoxical. The problem says that the paradox is deep. Here's why. Take the principle that a square is a four-sided equiangular equilateral figure. Anything that meets it is a square, everything else is not. A descriptive principle only applies to things that meet it. Now take the principle that drivers in America should drive on the right.³ Drivers subject to it only sometimes meet it. A normative principle applies to some things that do not meet it. In fact, a principle can be normative for a kind only if tokens can fail to meet it. The application conditions of descriptive and normative principles thus rule out a principle being descriptive of and normative for an activity.

Tie this objection to divisionism. The descriptive principle then picks out necessary properties of the kind that are part of its nature. A normative principle can only pick out *contingent properties* of a kind that tokens can have or lack. The content of these principles then differs. One principle cannot play both roles. The content necessarily differs, though, only if every token must have a property that is part of the nature of the kind. The problem, then, in effect just says that constitutivism is incompatible with necessitarianism.

The problem of bad action also assumes separatism about goodness, at least so long as multiple kinds can have good tokens and they can also have bad ones. This assumption is harder to expose, but doing so matters for the forthcoming argument. Bear with me. Here goes. So long as any kind that can have good tokens can also have bad ones, goodness is not part of the nature of the kind. While this claim

3. I put normative terminology in the principle in order to strengthen the appearance of paradox. I do not think that a normative principle needs normative terminology. Just think about how laws like 'Objects fall to the Earth at 9.81m/s^2 ' differ from laws like 'Motorists drive on the right side of the road in America'. We use normative terminology to explicate this distinction, but nothing goes missing without it.

is about the nature of the kind, it is also implicitly about the nature of goodness. Take equilaterality, equiangularity, and four-sidedness. Every square has these properties, but not everything with one of those properties is a square. Any regular polygon is equilateral and equiangular. Any quadrilateral is four-sided. Apply the principle behind necessitarianism to these properties and being a square is not part of the nature of any of them. Apply the principle behind necessitarianism to goodness and being the property of an action or anything else is likewise not part of the nature of goodness. An action is good, then, when it has a property whose nature is independent of the nature of our action. That just is separatism about goodness. Necessitarianism is in this way inseparable from separatism.

On the basis of the problem of bad action, divisionists reject constitutivism. I prefer to discard divisionism. I cannot argue for that move here. My aim instead is to present constitutivism on its own terms so that it can get a fair hearing. Such a presentation and hearing are possible, though, only while keeping in mind that constitutivism rejects the divisionist assumption that structures much practical philosophy and much discussion about constitutivism.

4. ESSENCE AND IDENTITY

I here will present the basic metaphysics of constitutivism and contrast it with divisionism. One aspect of this view, which I explain more fully at the end of the paper, is that constitutivism is not a view about all kinds. It is only about kinds whose tokens are subject to what I call internal normative standards. It is thereby compatible with the truth of necessitarianism with respect to other kinds. Let me thus first explain the difference between what I call *internal* and *external* normative standards so that I can explain the scope of the thesis after I articulate it.

A token is subject to an internal standard if the nature of its kind establishes that standard. It is subject to an external standard if something else establishes that standard. To establish a standard is to explain why it exists, not only explaining *that* it exists but also explaining *why* it is a normative standard. It is an explanatory metaphysical relationship, and explanation affords understanding. This is the metaphysical correlate of the familiar point that an account of the goodness of something must explain why it is worth choosing. With respect to internal standards, something is worth choosing as what it is. In the case of external standards, something is worthy of choice relative to the use, purpose, or whatever that establishes the standard in question.

Consider philosophy papers. Having interesting material on a relevant issue is an internal standard for them. A bad one might consist of a list of dramatic text messages. A good one might meet basic challenges in action theory and metaethics by unifying the nature of normativity and the activities of the living in a metaphysics of capacities. The nature of philosophy papers establishes this standard, and a paper is subject to this standard because it is a philosophy paper. Being useful as a doorstop, though, is an external standard for such a paper. It does not apply to one as such, but one can be better or worse for that use. That is why my reviewer insults me if she says ‘But it’s a good doorstop!’ after an otherwise negative review.

The heart of constitutivism is the idea that internal standards are possible because a property can be part of the nature of a kind even though tokens can lack it. The nature of the kind ranks its tokens according to whether they have those properties. When a token is good, the nature of its goodness depends on the nature of the kind in question. In this way, constitutivism differs from the divisionist view of kinds and goodness. Let me explain.

Think about glasses. Some are red, some blue. Redness is a contingent property of glasses and is not part of their nature. This set of claims is fine, but it does not move *from* the contingency of redness to glasses *to* its not being part of their nature. There is no ‘therefore’ there, to filch a phrase from Gertrude Stein. Now think about how even though only some glasses have intact lenses, whether one does is significant. A pair without them is to that extent a bad pair, whereas whether one is red is neither here nor there with respect to the internal standard of glasses. Why?

We create artifacts for purposes. The nature of an artificial kind is what a token needs in order to serve that purpose.⁴ Glasses have clear lenses, a mechanism that lets them to sit in front of our eyes, and other things that let them correct their intended wearer’s vision. Those claims are about the kind, not its tokens. Not every pair has these properties. Lenses fall out and get scratched, frames bend and break. Tokens of an artificial kind can lack some properties that are part of their nature.

Call a property that is part of the nature of a kind an *essential property* of that kind. Call any that tokens can have that is not part of the nature of the kind an *accidental property* of the kind. Contingent properties can then be either essential or accidental properties of an artificial kind. If the property serves the function of that kind, it is an essential property of the kind. If not, it is an accidental property of the kind.⁵ Hence, whereas necessitarianism says that only necessary properties of a kind can be essential properties of it, constitutivism says that contingent properties too can be essential properties of an

4. Philosophers sometimes use ‘artifactual kind’ instead of ‘artificial kind’ because ‘artificial kind’ might seem to contrast with ‘real kind’ instead of ‘natural kind’. I trust you, dear reader, to keep the actual meaning in mind.

5. What about the possibility of a bad artificial kind? On the one hand, an artificial kind that does not serve its purpose is a bad artificial kind. Perhaps most exercise sold on television at 3am is like this. On the other hand, since an artificial kind inherits its purpose from the activity in which we mean to use it, an artifact made for use in a bad activity is a bad artificial kind, at least until we reconstitute it for another purpose. Perhaps iron maidens were like this until we reconstituted them as education pieces about the depths of depravity to which human beings can descend. The internal standard for our actions thus limits the artifacts we can legitimately create. I say more about the internal standard for our actions in a bit.

artificial kind.

A token of an artificial kind is good to the extent that it has the contingent essential properties of that kind, bad to the extent that it lacks them. Goodness is thus not simply another property of the kind, as if we might find it glued to the lenses or screwed to the frame. It is a different kind of property. Tokens possess it, not kinds, and they possess it to the extent that they have their contingent essential properties. As we might put it, it is a *resultant property* of a token of an artificial kind. And this fact is no accident. 'Artificial kind' picks out a type of kind whose members establish internal standards for their tokens. What an artificial kind *is* determines what its tokens *should be* because of what an *artificial kind* is.

This relationship between an artificial kind and its tokens is isomorphic to a constitutivism view of capacities. Bundle up the essential properties of an artificial kind and you have a principle that describes its nature and is thereby normative for its tokens. Bundle up the essential properties of a capacity and you likewise have a principle that describes its nature and is thereby normative for its exercises. And this explanatory structure puts the lie to the problem of bad action. The principle plays its descriptive role with respect to the general thing, its normative role with respect to the particulars. Explicitly distinguish the general from the particular and the problem dissolves.⁶

Of course, we create artifacts for purposes, but no one creates our capacities, let alone for a purpose. Still, capacities are potentialities to do stuff. That is why they differ from each other given what

6. One reason to formulate constitutivism in terms of capacities rather than activities is that capacity predication forces you to make this distinction explicit. You must say 'the capacity to A' to talk about the general and 'an exercise of the capacity to A' to talk about the particular. You can make this distinction explicit with talk of activities by saying 'the activity of A-ing' and 'an instance of the activity of A-ing'. However, you need not. You can use 'A-ing' to talk about the general and the particular. Such language obscures a crucial aspect of the view. Similarly, although I do not discuss it in this paper, the distinction between my possessing a capacity and its state of development is essential to constitutivism. You can make this distinction in terms of activities by distinguishing my being able to engage in an activity from my competence at that activity. This way of marking the distinction is cumbersome, though, and less clear than distinguishing between possessing a capacity and its state of development.

they are potentialities to do. Just as the essential properties of an artifact are what a token needs in order to perform its function, so the essential properties of a capacity are what an exercise needs in order for the bearer to succeed in the activity. Goodness is thus a resultant property of exercises, not of capacities. Exercises are good to the extent that they have the contingent essential properties of the capacity, bad to the extent that they lack them. And this fact is no accident. The name 'capacity' picks out a type of potentiality whose members establish internal standards for their exercises. What a capacity *is* determines how an exercise of it *should go* because of what a *capacity* is.

There are thus four types of possible properties of kinds:

- (1) Necessary properties of the kind that are part of its nature.
- (2) Necessary properties of the kind that are not.
- (3) Contingent properties of the kind that are part of its nature.
- (4) Contingent properties of the kind that are not.

Whether necessitarianism or constitutivism is true about a kind depends on whether its nature includes contingent properties. If it does, its tokens are subject to an internal standard and have the resultant property of goodness to the extent that they possess those contingent essential properties. Goodness is a property of tokens, though, not of kinds. It thus is not within any of four types of possible properties of the kind.

Hence, a token is good to the extent that it has the essential properties of that artificial kind or capacity and thereby realizes its nature. It is bad to the extent that it lacks them and thereby does not fully realize its nature. Good tokens more fully exemplify their kind than bad ones, not because they have the property goodness but because they more fully possess the contingent essential properties of the

kind. They are thereby good. Goodness marks intrinsic perfection or excellence, badness privation or defect, of the token.

This metaphysics of goodness differentiates it from redness in a way that puts the lie to the analogy characteristic of separatism. The relationship between goodness and badness differs from the relationship between redness and blueness. For a token or exercise to be good is for it to possess the properties it by nature possesses. For it to be bad is for it to lack them. The goodness or badness of something is thus relative to the kind in question in a way that the redness or blueness of something is not. Constitutivism is thus incompatible with separatism. Moreover, badness is the absence of goodness or, more accurately, of the contingent essential properties that establish the internal standard for tokens of a kind. Blueness is not likewise the absence of redness or of any properties that might make an object red. Goodness thus has metaphysical and explanatory priority over badness in a way that redness does not have over blueness. The bad is not merely the non-good, which lies outside of the ambit of normativity. Badness is a failure within the house of goodness.

5. THE REALITY OF GOODNESS

Why is goodness not a contingent property that is not part of the nature of the kind? Some tokens possess it but not others, and they possess to the extent that they possess the contingent essential properties of the kind. Why, then, can only tokens possess goodness? After all, in some sense, only tokens can possess any property. Tokens are red or blue, not the kind itself, and yet nothing stops redness or blueness from being a contingent property of kinds, whether essential or not. Why is goodness different? In order to answer this question, let me lay out the familiar argument that the logical

features of the adjective 'good' rule out certain views of goodness. Discussing this argument reveals the significance of saying that goodness is a resultant property that is not among the four types of possible properties of kinds.

P.T. Geach's 'Good and Evil' is the first cause of recent arguments of this kind. He argues that 'good' is an attributive rather than predicative adjective and concludes that goodness is a relational rather than absolute property. Consider the predicate-form 'is a *A N*', where '*A*' stands for an adjective, '*N*' for a noun-phrase. '*A*' is a predicative adjective if 'is *A*' and 'is a *N*' are complete predicates. Take 'beige'. 'Murgle is a beige domestic cat' breaks up into the complete thoughts 'Murgle is beige' and 'Murgle is a domestic cat'. Moreover, the following inference is sound: Murgle is a beige domestic cat, he is a mammal, and therefore he is a beige mammal. Not so with attributive adjectives. Take 'big'. 'Murgle is a big domestic cat' does not break up into 'Murgle is big' and 'Murgle is a domestic cat'. If it did, the following inference would be sound: Murgle is a big domestic cat, he is a mammal, and therefore he is a big mammal. This inference is unsound, though, thank goodness for his sake. Domestic cats, even the largest, are small mammals. Just compare them to elephants and whales. Attributive adjectives thus do not detach from the noun phrase in a complete predicate because the noun provides the standard for the adjective. They have a gap in them that a class-fixing kind term must fill in order to complete predicate. In this way, '*X is big*' is not a complete thought. And 'good' functions like 'big'. 'JDF is a good karaoke partner' does not break up into 'JDF is good' and 'JDF is a karaoke partner'. Inferences of the form 'JDF is a good karaoke partner', 'JDF is a *N*', and therefore 'JDF is a good *N*' for any *N* that applies to me are unsound. Otherwise, to be good at something entails being good at everything I do, and likewise to be a good something entails being a good everything for every kind term that applies to me. That is absurd.

My talent at video games does not translate so easily to philosophy, alas. 'Good' differs from 'red', and goodness differs from redness.

I accept this argument as far as it goes. I will not discuss criticisms of it, with one exception later. I shall, though, explain a division within the class of attributive adjectives that is important for the metaphysics of goodness. Adjectives like 'small' and 'early' relate a token to other tokens of the kind. They do not relate it to the kind itself. 'Small' picks out a property of comparative size among objects in the class. 'Early' picks out a temporal relationship of some events to others in a class on a time line. Adjectives like 'pure' and 'good', though, pick out a relationship between the kind and its tokens, not between the tokens. A token is pure gold, for example, if it is nothing but gold. Although this fact implies certain relationships to other tokens of this kind given their purity, it is not about those relationships. Purity instead belongs to tokens given their relationship to the kind. Likewise for 'good' on my metaphysics of goodness. An action is good to the extent that it has the contingent essential properties of the capacity in question. Although this fact implies certain relationships to other exercises of this capacity given their goodness, it is not about those relationships. Goodness instead belongs to tokens given their relationship to the kind.

Attributive adjectives thus divide into at least two subclasses, *gradable* ones and *archetypical* ones. 'Small' and 'early' are gradable, 'pure' and 'good' archetypical. To test to which subclass an adjective belongs, ask whether it can apply to every token of the kind, past, present, and future. If it can, it is archetypical. If it cannot, gradable. Nothing precludes every lump of gold being pure or every human action being good. In contrast, only some elephants can be small, only some admission deadlines early. 'Small' picks out a contextually fixed threshold along an independently intelligible spectrum of tokens

arranged according to size. The spectrum is prior to this threshold, and ‘smaller’ picks the placement of one token relative to another on this spectrum. In contrast, ‘pure’ and ‘good’, at least on my metaphysics of goodness, do not pick out contextually fixed thresholds along independently intelligible spectra.⁷ ‘Pure’ and ‘good’ mark the end point of spectra in light of which everything else is to some degree impure and to some degree bad.

Goodness, like purity, cannot be a contingent property of the kind because it concerns the relationship between tokens and the kind. To treat it as a contingent property of the kind that is not part of its nature does not capture the characteristic of its relationality that unites it with purity and divides it from tallness and smallness. Goodness just is the property of a token having the contingent properties of the kind that are part of the nature of the kind. A token can be good because the kind has contingent essential properties. A token is good when it has those. To call goodness or badness a resultant property of tokens of a kind highlights this aspect of the property.⁸

6. ACTION WITH DETERMINATION

I turn to the basic challenge of metaethics in the next section. I first shall discuss a common objection to Geach’s grammatical argument and explain how it fits with metaphysics of capacities. This response also lets me respond to the ‘shmagency’ objection to constitutivism.

One issue with Geach’s argument is that it is hard to understand how it shows that we can be

7. At least, they do not just mark on those thresholds. We can use these words that way, but it is not their basic role, at least with respect to understanding the properties in question.

8. Were there time and talent, I would connect this discussion to general issues about properties and predication. The contemporary view treats predication as flat in the sense that all predicates pick out properties that are in an important sense of one kind. The classical view distinguishes between types of predication, not only separating substance from accident predication but also distinguishing predicates that pick out an object’s properties from other true things that can be said about an object. I am sympathetic with this view, but I cannot explain and defend it here. I am instead defending an account of goodness that is most at home in the classical view while operating largely within the confines of the contemporary view.

subject to moral or prudential requirements in our action as such. After all, ‘thief’, ‘swindler’, ‘mob boss’, and ‘capitalist’ are all class-fixing kind terms that seem to allow for good and bad tokens. Or, talking about characteristic act-types of these chaps, stealing, swindling, protecting the family above all, and privatizing ownership of the means of production are all act-types that seem to allow for successful and unsuccessful instances.⁹ Assuming for the sake of argument that human beings are subject to prudential and moral requirements on our action as such that preclude acting in these ways, what grammatical features can show that human beings cannot legitimately get up to this stuff?

Since my argument for constitutivism is primarily metaphysical, not grammatical, the problem needs a different twist. The change is easy enough, though stating it just right is a bit tricky. If we are subject to rational, prudential, and moral requirements in our action as such, we lack a capacity to act in irrational, imprudent, and immoral ways. That is a bold claim given the manifest reality of irrationality, imprudence, and immorality. How does the metaphysics of capacities explain the possibility of such standards? What is the same thing, how does this metaphysics show that a being can have a capacity to act morally, prudentially, and rationally and thereby lack a capacity to act immorally, imprudently, and irrationally.

The key is the nesting determinable-determinate structure of capacities. Take my capacity to walk forward, itself a determinate of my capacity to walk, itself a determinate of my capacity to act. If to walk forward is in part to put one foot in front of the other, any determinate of it must include this fact in

9. I here assume that all constitutivists mean to establish *some* substantive requirements on our action, where ‘substantive requirements’ means requirements more robust than the instrumental requirement. I do not think that constitutivists must show that these requirements are prudential and moral. As far as the metaphysics of capacities says, they might be something else. I am using prudential and moral requirements because they are familiar. If you do not like them, pick something else and alter the act-types accordingly.

its principle. It can modify this principle in various ways—‘put one foot in front of the other at a leisurely pace’ is a different determinate than ‘put one foot in front of the other cautiously’. But it must stay true to the determinable. This is why strutting, slinking, and sauntering are determinates of my capacity to walk but walking while remaining completely still or walking without artificial aid in a zero gravity environment are not. These phrases try to describe an activity that is internally contradictory in one way or another and hence is not a real activity, or at least not a human one. We cannot have a capacity to act in these ways. To set off to so act is to automatically fail in the way that to set off to square the circle using only compass and straightedge is to automatically fail. It is not that I am not exercising my determinable capacity to act. It is instead that I am exercising it in such a way that success is impossible because I in principle lack the capacity to act in this way. Only certain determinates of a determinable capacity are possible.

The principle of our capacity to act likewise restricts its determinates. If this principle includes moral, prudential, and rational content, determinates of it obey moral, prudential, and rational requirements. A capacity to act immorally, imprudently, or irrationally would be as impossible for us as a capacity to walk forward backward. The phrases ‘act immorally’, ‘act imprudently’, and ‘act irrationally’ would try to describe activities that are internally contradictory in one way or another and hence are not real activities, or at least not human one. We cannot have a capacity to act in these ways. To set off to so act is to automatically fail in the way that to set off to sail to the edge of the earth is to automatically fail. Again, it is not that I am not exercising my determinable capacity to act. It is instead that I am exercising it in such a way that success is impossible because I in principle lack the capacity to act in this way. Just as I can be on the boat sailing while not possibly succeeding in sailing off the edge of the earth, so I can be

disregarding necessary means and undermining my good and yours as well while not possibly succeeding in acting irrationally, imprudently, or immorally. While I do not here argue that the principle of our capacity to act in fact establishes these requirements, the metaphysics of capacities establishes the possibility of such requirements on our action in a way that grammar of 'good' does not.

I suspect that an air of mystery might surround this explanation. I cannot say everything that I need to say to dissolve it.¹⁰ It might help, though, to distinguish this metaphysics of constitutivism from a certain epistemological assumption often associated with it. Many critics and defenders of constitutivism assume that any normative standard that governs our action as such must be analytically derivable from the concept of self-conscious or rational action. This epistemological assumption does not follow from the metaphysics. The only epistemological implication of the metaphysics is an inferential link between knowledge of the nature of a capacity and knowledge of the standards that govern its exercises. You can come to know either by inferring it from the other. If you know the nature of a capacity, you know its principle in its descriptive role. If you know the standard that governs its exercises as such, you know that principle in its normative role. Since one principle plays these roles, you can infer either from the other. Still, although knowing the nature of a capacity is enough to thereby know the standard that governs its exercises as such, this metaphysics says nothing about *how* you come to know the nature of the capacity. It does not say that the nature of every capacity is analytically derivable from the concept of the activity that differentiates one capacity from another. Knowledge of the nature of a capacity might be synthetic in the way that knowledge of the nature of water or of human digestion is synthetic. Hence, as far as the

10. Such a discussion would have to work through how a capacities divides its exercises into the good and the bad and thus how it divides what is possible with respect to it into two sub-classes. I take up general issues about capacities and possibility in "Practical Possibility".

metaphysics says, the normative standard that governs exercises of a capacity as such need not be analytically derivable from the concept of the activity that differentiates it from others. With the possibility of this synthetic knowledge, though, comes the possibility of actions that are impossible for us even if such impossibility is not obvious. That is how it is if irrational, imprudent, and immoral actions are incompatible with the principle of our will. We cannot successfully digest various things even though analysis of the concept of human digestion does not disclose such impossibility. Just so, we cannot successfully act irrationally, imprudently, and immorally if the principle of our capacity to act includes rational, prudential, and moral content even though analysis of the concept of rational action or human action does not disclose such impossibility.

This determinable-determinate structure, by the by, puts the lie to David Enoch's 'shmagency' objection, which says that constitutivism cannot establish the authority of internal standards for our action even granting its view of our action. His criticism turns on the claim that a question like 'Why be an agent who performs actions and not a shmagent who performs shmactions, which are like actions except that they are not subject to those standards?' is well-formed even if constitutivism is true. According to this metaphysics of capacities, though, this question is ill-formed. In order to do the needed argumentative work, the principle of shmacting must conflict with the principle of our capacity to act. Since we by nature have a capacity to act, a capacity to shmact would be a determinate of that capacity, in conflict with its determinable. 'Act shmactingly' would then try to describe an activity that is internally contradictory and hence not a real activity, or at least not a human one. We cannot have a capacity to act in this way anymore than we can have a capacity to drink water while not drinking H₂O. To set off to shmact is to automatically fail in the same way that trying to drink water but not H₂O is to

automatically fail. The shmagency question is no more intelligible than the question ‘Why walk forward instead of forward backward?’, though the unintelligibility is not so obvious. Enoch thus cannot grant the metaphysics and run his criticism. He must deny the metaphysics. The shmagency objection is useless for that task.

7. THE SOURCES OF NORMATIVITY

Let me address the basic challenge of metaethics and explain the different kinds of normative standards, their scope, and the relationships between them. Tokens of artificial kinds and exercises of capacities are subject to internal standards. However, these standards differ in a key respect. Internal standards of artifacts are derivative. Internal standards of capacities are not. We create and use artifacts for a purpose, which determines the essential properties of the kind. An artificial kind thus derives its essential properties from its role in the activities in which we use it. Since to be active is to exercise a capacity, the principle of an artificial kind and the internal standard for its tokens derives from the nature of some of our capacities.

Internal standards of artifacts are in this respect like the external standards to which anything can be subject. We the living use things in many ways. This use constitutes an external standard for something. My cat might be a bad pillow for me, a lighter might be a bad toy for him, and a rock might be bad for him to headbutt. In each case, something is subject to a standard given the use that a being makes of it. Although we make use of these things, though, we do not make them for our use. These standards are thereby external standards. They apply to something because of how we use it but not because of what it is.

There are thus three kinds of standards: primitive internal standards, derivative internal standards, and derivative external standards. Non-living non-artificial stuff is subject to derivative external standards. Artifacts are subject to derivative internal and external standards. Living beings are subject to derivative external standards and to primitive internal standards in exercises of capacities. The capacities of the living ground every standard. They are the sources of all normativity.¹¹

8. THE GOODNESS OF REALITY

Although that explanation meets some aspects of the basic challenge of metaethics, it does not explain the unique role that the standard for our capacity to act can play in its exercises. No does it address the basic challenge of action theory. To meet these challenges, I shall revise the problem of bad action and

11. A full discussion of these issues would explain how internal and external standards interact. I cannot engage this issue in detail here, but let me explain three areas where it is relevant and how it applies. I hope to take up these issues in future work.

First, although my focus is on capacities, artifacts, and using stuff, much normative thought about the living focuses on normative standards for their parts—for their organs, for their limbs, and so on. I take the internal normative standards for these parts to derive from the normative standards for the characteristic activities in which the being uses those parts. A broken leg is defective, say, because I cannot walk well or at all with it. However, these internal standards for the parts do not imply that I am in error when I use that part for some other purpose in the exercise of some other determinate of my capacity to act. To take what is likely to be the first example to pop up, although the internal standard for a human penis concerns whether it is fit to expel liquid waste and whether it can play its role in sexual reproduction, I am free to use this organ in other kinds of activities like non-reproductive sexual activity without defect or error. So long as I do not use the organ in such a way as to damage its ability to perform its function, nothing in the metaphysics says that I am to *only* use an organ in a way that meets its internal normative, anymore than the metaphysics says that I am to only use an artifact in a way that comports with its internal standard.

Second, I take this distinction between kinds of normative standards to explain the in my opinion unfortunately named *wrong kind of reasons* problem. Briefly, the problem is that considerations that bear on whether to exercise a capacity seem to fall into two classes. When it comes to an exercise of my capacity for belief, say, evidence that bears on the truth of something seem to differ from financial inducements to believe it. Likewise for our capacities to desire, admire, and so on. The question is about what explains this difference. On my metaphysics, this difference just is a difference between internal and external standards for exercises of a capacity.

Finally, this distinction explains the truth behind evidentialism. Roughly, evidentialists think that truth-related considerations are relevant to exercises of theoretical reason in a way that, say, whether happiness-related considerations are not. Unfortunately, many of them formulate this idea in terms of the *impossibility* of ‘practical reasons for belief’, thereby taking on the unenviable task of explaining how this view accommodates self-deception, wishful thinking, and epistemic akrasia. They instead should say that the nature of our capacity for theoretical reason establishes truth-related considerations as the internal standard for its exercises. Practical considerations are at best external standards. Self-deception, wishful thinking, and epistemic akrasia are instances of an agent incorrectly exercising her capacity in accordance with external standards rather than internal standards. Whether there is ever an instance when an agent can correctly exercise this capacity on the basis of considerations that are external standards for it is a different question that I do not here address.

show that the way to answer it highlights aspects of exercises our capacity to act that address the remaining challenges.

The revision of the problem turns on thinking of exercises of our capacity to act as events, not processes. An event is a fully determinate particular. In contrast, a process acquires intrinsic properties as it progresses towards a terminus. Consider the chap in a Peter Railton vignette who says that

there simply is no question of why Railton spelled ‘correspondence’ with an ‘e’. Spelling is a normative concept—acts of spelling constitutively involve satisfying the norms of spelling. So he couldn’t have spelled the word with an ‘a’—to have written ‘correspondance’ wouldn’t have counted as a spelling of ‘correspondence’ at all. (Railton [1999] 323-4)

If you stare at the end result or event, the word is either spelled correctly or not. There is nothing else to say, at least so long as Railton is not chalk in hand at the ready at the board. An incorrect spelling, though, is no spelling of the word. There is then no event of spelling ‘correspondence’. How can there be an event of a failed spelling of ‘correspondence’, though, if there is no event of spelling it, anymore than there can be an event of spelling it quickly without one of spelling it? As with spelling, so with every action.¹²

Exercises of our capacity to act are fundamentally processes, though, not events. When I exercise my capacity to act, I am acting for as long as I am exercising it. There is an event of my action only when I am done acting. Say I represent building a monumental sandcastle. I head to the beach and start dreaming, designing, digging, dragging, and developing. These things eventually are parts of the whole that is my action. While I am acting, though, they are phases in an ongoing process. Unlike an action that

12. The details of this problem and response are specific to capacities whose actualizations are processes. The actualization of some capacities, though, are states and others are activities, where a process differs from an activity because it has an internal terminus. Those capacities face a version of the problem because it derives from the fact that what describes the nature of a capacity also describes its perfect or complete actualization. The issue is about how an imperfect actualization is possible if the nature of the capacity relates in this way to its perfect actualization. A general response would turn only on how a capacity by nature divides its exercises into the good and the bad.

is a 'thing done', acting is a 'doing of a thing'. While I am acting, the action is incomplete, ongoing, and to that extent indeterminate. I have made the walls, the bastions, and the turrets. The steeple and the palace and all that jazz, though, are still to come. Soon my sandcastle will be complete. Here I go! My action gets more determinate as I act, and I act given my grasp of what I am doing and why. When I stop by finishing the sandcastle, the process of my acting constitutes the event that is my successful action. I built that sandcastle! When I otherwise stop, the process of my acting constitutes the event of my failed action. A wave washes over us, leaving a crab and my broken dreams floating inside a puddle of mud. In the end, then, it is fine to say that actions are events. It is fine only literally in the end, though, because there is an event only after the process ends, an action only when I am no longer acting.

Because the exercise of this capacity is primarily a process, its principle applies primarily to my acting, not my action. And the fact that I can be in the middle of something without being finished is boring. That just is what it is to be in the middle of a process. It takes place over time and can be interrupted, even terminally. The principle of the capacity applies to unsuccessful exercises because it primarily applies to them *as they are ongoing*. It applies to the event of my action because it applies to the process of my acting that eventually constitutes that event. That is why writing 'correspondance' is a misspelling of 'correspondence'.¹³ The revised problem is as much an illusion as the original.

As with our capacity to act, so with other capacities. My cat is crossing the street for a while. If all goes well, it ends when he has crossed it. If all does not go well, when it is over, he was crossing it but did

13. How far can I fall from progressing towards the end and still be acting? There is nothing straightforward to say. We perform actions in many ways, many of which have gaps and other things that do not foretell success. Think about the false starts, restarts, dead ends, flailing, and everything else that does not advance the writing of a paper and may even hinder it. Yet the process ultimately constitutes an event that includes it all, successful or not. Tolstoy, or at least his action-theoretic counterpart, is only half right. Success, like failure, comes in as many ways as there are actions.

not cross it. Just so, I am digesting fruit for a while. If all goes well, it ends when I have digested it. If all does not go well, when it is over, I was digesting it but did not digest it. The process eventually constitutes an event of a complete or incomplete process depending on whether it reaches its terminus. The event exists because the process exists and is derivatively subject to normative standards. The same explanatory structure applies. Differences between cases depend on differences between the capacities exercised.

This shared explanatory structure leaves room for the difference between our capacity to act and other capacities. Constitutivism thereby meets the basic challenge of action theory. All capacities and their exercises have the same metaphysics and are subject to the same form of explanation. Our action fits into a metaphysics that applies to anything the living does. This metaphysics distinguishes these things from stuff that just happens while leaving room for different kinds of capacities. Some things can be done on the basis of our understanding of the principle of that kind of thing. They are self-conscious, and exercises of these capacities are thereby unique even though their bearers live and exercise them in the same world as everything else exists and happens. Because we have a self-conscious capacity to act, we can act from our understanding of the principle that describes its nature and is thereby normative for its exercises. We can guide ourselves by the internal standard for what we are doing. In this way, the metaphysical uniqueness of our capacity to act explains its normative uniqueness and meets the left over aspect of the basic challenge of metaethics.

Like necessitarianism, the constitutivist metaphysics links something general with its particulars. Assuming for the sake of explanation the truth of constitutivism, the error in divisionism is to try to assimilate distinct types of kinds to one metaphysics. Applying constitutivism to every type of kind

would commit the same error in reverse. With respect to geometrical kinds, a principle describes the nature of a kind in a way that divides things into tokens and non-tokens. It does not establish a class of defective tokens. All of the essential properties of the kind are necessary properties of it. Necessitarianism is correct about these kinds, as well as any other whose tokens are not subject to an internal standard. It is not true of all kinds. In this way, the genus *kind* has at least two species, which transpose the explanatory structure that links something general with its particulars into different keys that license different concepts. Normative concepts have a place with respect to constitutivist kinds that they lack with respect to necessitarian ones. They have a place when some of the essential properties of the kind are contingent properties of it. In recognizing the diversity of types of kinds, we recognize the diversity of the world.

Badness, then, is primarily the intrinsic defectiveness of an exercise of a capacity of the living. It is derivatively the intrinsic defectiveness of an artifact and the extrinsic defectiveness of something for the use that a living being makes of it. It is an error in the house of the living, of their nature, their artifacts, and their uses of things. Goodness is primarily the intrinsic perfection of an exercise of a capacity of the living, a realization of its nature. It is derivatively the intrinsic perfection of an artifact and the extrinsic perfection of something given how a living being uses it. It is a part of the reality of the constitutivist kinds, of things by nature subject to normative standards. The Real is the Good, not because everything real is good and not because everything good is real but because some things are by nature subject to a standard of goodness. The Good is in this way neither more nor less mysterious, neither more nor less unnatural, than the activity of the living. It is as ordinary and real as life itself.

PRACTICAL POSSIBILITY

Further, matter exists in a potential state, just because it may attain to its form; and when it exists *actually*, then it is in its form. And the same holds good in cases in which the end is a movement, as well as in all others. ... And while in some cases the exercise is the ultimate thing (e.g. in sight the ultimate thing is seeing, and no other product besides this results from sight), but from some things a product follows (e.g. from the art of building there results a house as well as the act of building), yet none the less the act is in the former case the end and in the latter more of an end than the mere potentiality is. For the act of building is the thing that is being built, and comes to be—and is—at the same time the house. (Aristotle *Metaphysics* Θ 15-30)

1. 'CAN' AND CONDITIONS

A characteristic 'can' often pops up in practical philosophy:

The Success Condition A principle governs my action only if I *can* comply with it.

The Error Condition A principle governs my action only if I *can* deviate from it.

The Practical Cognition Condition A principle governs my action only if I *can* self-consciously comply with it—that is, only if I *can* act as it requires or permits and refrain from acting as it prohibits from my understanding of what it requires, permits, and prohibits.

The Alternate Possibility Condition I am responsible for acting in some way only if I *can* do otherwise.¹⁴

You can multiply examples if you like. These conditions attract and repel us, at least in part because of the modal vocabulary in them. Take the success and error conditions. They ring true because they apply general conditions on normative principles to the case of our action. Any normative principle divides

14. A few notes on terminology. First, I call these theses *conditions* in order to avoid ambiguity when I discuss *conditions* on *principles* governing our action. Second, I prefer 'The Success Condition' to 'Ought Implies Can' for three reasons: (1) it suggests the relationship with the error condition that I later explain; (2) it indicates that this condition is an instance of a general condition on normative principles for any activity of any being. Some philosophers might deny 'ought' a place where they allow talk of success—'I successfully digested my food' and 'my cat succeeded in tipping over the trashcan' are I hope fine to ears stopped up against an 'ought' with respect to human digestion or feline action; (3) 'implies' is inapt because the condition is about a metaphysical relationship, not a logical relationship. If ought-propositions imply can-propositions, they do so because the success condition is true. Third, I do not qualify 'responsible' with 'morally' in the alternate possibility condition because the issue of interest is not limited to responsibility for complying with or deviating from moral requirements as opposed to, say, prudential and rational requirements. I trust readers to distinguish this topic from causal responsibility, legal responsibility, and the like without an explicit qualifier.

what falls under it into the compliant and deviant. Just think about principles that do not, such as ‘Do A or do not do A’ and ‘Do A and do not do A’. I cannot deviate from the first or comply with the second. Nothing about action in general or our action in particular is the source of the issue. Principles with such form cannot govern our digestion anymore than they can govern our action. Compliance and deviance, though, are a matched pair. There is no normativity if you cannot go right, none if you cannot go wrong. A principle governs a kind only if tokens of that kind can meet and fail to meet it, as these conditions codify with respect to our action.

Stare at the conditions, though, and the picture that they paint goes out of focus. Without the success and practical cognition conditions, principles that govern our action would be insensitive to our abilities. Try to imagine moral, prudential, or rational requirements with nothing to do with our abilities and these conditions will attract you.¹⁵ Descriptions of abilities, though, can be very fine-grained. At some point, they would rule out so much that it is unclear how I can be subject to moral, prudential, or rational requirements in various situations in which my immoral action seems to be a failure. Ignoring and belittling my partner is wrong and my action a failure even if I cannot bring myself to stop. Even more fine-grained abilities would rule out even more principles from governing our action, to a point that it becomes unclear how a principle conditioned by them can say what I must, may, and must not do rather than say merely what I will do. How can principles that govern my action depend on my abilities without losing their normativity? The normative character of these principles thereby attracts and repels us from these conditions.

15. I use ‘rational requirements’ to pick out things like the instrumental principle, consistency principles, and the continence principle. As will be clear, I do not think that moral and prudential requirements are arational in the sense that the ground of their authority is external to practical reason. Unfortunately, we lack a better term for the narrower set of requirements.

The error condition also seems to follow from the normative character of principles that govern our action. Think about how gravity differs from morality, prudence, or rationality. To ‘violate’ gravity is to falsify it, and it forces us to reformulate the principle. Not so with respect to morality, prudence, or rationality. Violations do not thereby disqualify a formulation of the principle. A normative principle allows for deviation. Yet this condition might seem to bring bad baggage. Douglas Lavin, for instance, argues that one interpretation implies that answers to practical questions are arbitrary, including answers to questions about whether to follow principles.¹⁶ The authority of moral, prudential, and rational requirements, though, cannot depend on an arbitrary choice about whether to follow them. Their objectivity rules it out. Any such condition on them would refute them. The normative character of principles that govern our action attracts and repels us from this condition as well.

Unlike the others, the alternate possibility condition is not about what governs human action. It instead conditions responsibility for complying with or deviating from those principles. Yet it likewise attracts and repels us because of the normative character of those principles. As we might put the basic idea, how I act must be up to me if I am responsible for complying with or deviating from a principle. But ‘up to me’ in what sense? As with the practical cognition condition, descriptions of abilities can differ substantially. Too permissive a description and I can be responsible for failing to meet a normative principle even though I have no hope here and now of meeting it. Too restrictive and I cannot be responsible for failing to meet a normative principle because of accidents of circumstance and situation that seem normatively irrelevant. What explains which notion of possibility gets the extension correct?

A characteristic ‘can’ also often pops up in objections to theses about our agency:

16. See (Lavin [2004] 446-9). I later discuss this argument in detail.

The Guise of The Good To act is to act in a way that I take to be good.

Objection: I *can* act in a way that I do not take to be good—here I take your pear for the sake of the wicked, and here I bite my nails indifferently.

The Action-Knowledge Link To act intentionally is to act knowingly

Objection: I *can* act intentionally without knowing what I am doing or why I am doing it—here I sign ten carbon copies in order to legitimize the contract even though I doubt that I can sign them all no matter how hard I push with my pen, and here I say harsh words in order to offend because of a repressed desire to hurt you.

Practical Reason Is The Will To judge practically is to decide.

Objection: I *can* judge that I should do something without deciding to do it, and I *can* decide to do something without judging that I should do it—here I judge that I should work on my paper while deciding to hit the club.

Again, you can multiply examples if you like. These objections might seem odd. After all, as formulated, the theses lack modal vocabulary. The objections are relevant, though, if the theses are claims meant to hold with necessity:

The Guise of The Good (Modalized) Necessarily, a human being acts in some way only if she takes acting in that way to be good.

Objection: Possibly, a human being acts in a way that she does not take to be good.

The Action-Knowledge Link (Modalized) Necessarily, a human being acts intentionally if and only if she acts knowingly.

Objection: Possibly, a human being acts intentionally without knowing what she is doing, and possibly, a human being acts intentionally without knowing why she is doing it.

Practical Reason is The Will (Modalized) Necessarily, a human being judges that she should ϕ if and only if she decides to ϕ .

Objection: Possibly, a human being judges that she should ϕ without deciding to ϕ , and possibly, a human being decides to ψ without judging that she should ψ .

If the objections are true, the modal interpretations of the theses are false.

In this chapter, I shall argue that a constitutivist metaphysics of capacities explains the correct interpretations of these conditions of agency and theses about our agency. The basis of this metaphysics

is the idea that a capacity is such that a single principle describes its nature and is thereby normative for its development and exercises. This metaphysics provides a *structured* view of possibility that divides what *is possible* with respect to a capacity into the sub-classes of what *I can do* because I have it and what *can happen* because I have it. I shall argue that this distinction and the one between possessing a capacity and its state of development explain the distinct notions of possibility in the conditions of agency. This metaphysics thereby explains the unity and diversity of the notions of possibility. I shall also argue that the unqualified nature of the theses does not show that they are claims about exercises of our capacity to act that are meant to hold with necessity. They are instead about the nature of our capacity to act and thereby are normative for its exercises. Since normative principles allow for the possibility of error, the theses allow for deviation from them. Far from being refuted by deviant exercises, I will explain why deviant exercises are actually evidence for them.

This paper is long, this story complicated. I argue in “The House of Goodness” that this metaphysics explains the possibility of normativity and shows that the capacities of living beings are the sources of all normativity. I here mean to exhibit some of the explanatory work that it can do. At the same time, I mean to reveal the relationships between these conditions and theses. Yet the prospect of taking them up is daunting. Each is an established topic of investigation, as are the relationships between some of them. I cannot engage any of these debates, let alone all of them. Each discussion is a rabbit hole. Even one trip down would detract from the narrative essential to understanding how they fit into an account of our agency. Only in this way can we understand any of them, or so I think. Yet to discuss these conditions and theses within the constitutivist metaphysical framework gives them a particular shape. They might look very different in another framework. Since I elsewhere argue that the

metaphysics is correct, I think that the shape is true. Still, the only way to defend this strategy is to give the account and show that it elucidates them and thereby helps us understand our agency. Even with respect to that task, no discussion here is complete. I can only provide the framework and outline how to approach things in light of it. We need a full picture of the world and our agency in it as these conditions make it out to be if we are to determine whether this world, this agency, is ours. A partial account of them is but a first few brush strokes in this painting, though indispensable indeed.¹⁷

2. THE POSSIBILITY OF PRACTICAL COGNITION

Whereas the success and error conditions apply a condition on normative principles in general to the case of our action, nothing like the practical cognition condition applies to normative principles generally. I can digest well or badly. I can understand human digestion. I can act from that understanding by eating healthy stuff. Yet I cannot digest from my understanding of human digestion. Similarly, my cat can act well or badly. He cannot understand feline action, though, let alone act from his understanding of it. Still, this condition rings true because it modulates the success condition into the key of self-conscious agency. We can act self-consciously, and I can come to understand this capacity through training, exercising it, reflecting on it, and every other part of an education. I thereby can act in

17. I also will not address general metaphysical issues about modality. Let me explain why. There are, I take it, at least two choice points with respect to the relationship between capacities or potentialities generally and other modal notions. The first is about which modal notion is most closely related to potentialities. Should we primarily link them with counterfactuals, say, or with possibilities? The second is about the nature of that relationship. Does something have a potentiality because a relevant counterfactual is true or because something is possible, or is the counterfactual true or is something possible because the object has the potentiality? Is there any explanatory priority here? Nothing that I say here turns on specific answers to these questions. With respect to the first question, on any answer to it, capacities have some relationship to possibility, whether direct or indirect. That is all my view requires. With respect to the second question, my concern is with the distinctions needed in order to understand the conditions and theses of interest. So long as you make these distinctions, the order of explanation does not matter. My rhetoric no doubt betrays sympathy with a view that takes potentialities as basic and relates them to possibilities directly. Still, nothing I say requires or establishes that framework. At best, I only show that these ideas are most easily and clearly expressed in it, and even then only by exemplifying, not arguing. I owe a great debt to Barbara Vetter's *Potentiality* with respect to these issues.

some way because I understand what the principle that governs my action requires, permits, or prohibits. Because human beings can act from thought about action, a principle governs my action only if I can act from my understanding of it. The practical cognition condition thus specifies the nature of compliance with respect to exercises of our capacity to act. It makes explicit what the success condition leaves implicit.

Inquiry into the conditions of agency should start with this condition because, as I shall explain, the notions of possibility in the error and alternate possibility conditions depend on the notion in it. But what notion does it use? Without an answer, the condition is indeterminate. Answering is hard, though, because of the *objectivity* and *internality* of the basic normative standard for our action. Let me explain these ideas before I link them with the issue about possibility.

We can ask at least two questions about the basic normative standard for our action. We can ask about its content, about what it requires, permits, and prohibits. We can also ask about its authority, about whether and why it governs our action. Why it and not another principle? Why does it govern us but not the other known animals, and why in action but not other things we do? An answer to the first question tells us what we must, may, and must not do, an answer to the second why we must, may, and must not do it.

Although we can ask these questions separately, their answers are not isolated. Even if I know what a principle *would* require, permit, and prohibit, it *actually* requires, permits, and prohibits me to act in certain ways only if it governs my action. Only then is it a normative standard for me. The answer to the second question is thus the holy grail of practical philosophy. It promises self-knowledge of our

capacity to act that accounts for the principle that governs us in action as such.¹⁸ An account of the conditions of agency is part of this self-knowledge. It partially explains the possibility of moral, prudential, and rational requirements on our action as such by establishing the reality of conditions on their possibility.

An answer to this second question must explain two aspects of the basic normative standard for exercises of our capacity to act. On the one hand, this principle establishes an objective normative standard for them. Moral, prudential, and rational requirements govern us independently of the idiosyncrasies of our psychologies. Just reflect on being subject to them and you will recognize this objectivity. On the other hand, this standard is internal to us. These requirements are not alien to us. They do not force themselves upon us from outside. We instead can recognize ourselves in them. If we could not, they would be impositions on us. Our apparent experience of their objectivity would be a misleading effect of their arbitrary influence on us.

An account of the principle of our capacity to act must explain this internality and objectivity. As with just about anything with two aspects, though, trying to explain one seems to muck up explaining the other. Take the practical cognition condition. It allows for a host of interpretations. The crucial ambiguity lies in the 'I can comply' clause. One challenge is to figure out the proper notion of possibility such that the principle can establish an internal objective normative standard for our action as such. Another is to explain why that notion is the right one. Only then do we understand this condition and hence start to understand our agency. I shall discuss the first challenge in the rest of this section. In the

18. This task mirrors Kant's claim that "reason should take on anew the most difficult of all its tasks, namely, that of self-knowledge, and to institute a court of justice, by which reason may secure its rightful claims while dismissing all its groundless pretensions" (Kant [1781/7] Axi). The task of critical philosophy is to come to reflective knowledge of the theoretical, practical, and reflective uses of reason. Self-knowledge of our capacity to act is reflective knowledge of that practical use.

next, I present two challenges to meeting the second.

What is the correct interpretation of this condition? What notion of possibility does it use? Logical possibility is too permissive.¹⁹ While I cannot be required to shut my eyes and not shut them, I also cannot be required to curl all of my digits and flex at least one toe. This action is logically but not conceptually possible because the contradiction is in the substance of the description of the action, not its form. A principle that governs our action cannot require such things. Conceptual possibility, though, is also too permissive. Some conceptually possible actions are metaphysically impossible, like drinking water and not drinking H₂O. The condition excludes principles that require such actions.

Physically impossible actions likewise fall outside the relevant sense of what I can do. I cannot be required to let myself free fall to Earth at a rate of acceleration greater than 9.83 m/s². Not everything that is physically possible for some being, though, is physically possible for us. I cannot be required to swim to the bottom of the Mariana Trench without mechanical aid. Human beings cannot survive under the sea. The pressure is too high, and we breathe air, not water. To confuse them is a danger. Since for fish it is the other way around, a suitable self-conscious fish can be required to act in this way. Still, the question 'Does this principle govern?' is incomplete until which beings it might govern in what activity is specified. When the question is 'Does this principle govern my action?', it is about me, who am not a fish, suitable or otherwise. With respect to human beings, this action is physically impossible, and a principle that governs our action cannot require such things. The notion of possibility in the condition is at least as restrictive as physical possibility relative to the kind of being subject to the principle.

There is more to us than what we can physically do, though, because we act from thought. I

19. I here indulge philosophers who distinguish logical, conceptual, and metaphysical possibility. If you do not like these distinctions, ignore them.

cannot be required to tell my partner that I love him in my sleep even though those sounds can exit my mouth while I sleep, perhaps even because of my thought about the goodness of such an event. Causation is weird. Likewise, I cannot be required to accidentally trip while I think about the circus even though I might end up on the ground while elephants and lions parade in my head. I can think about these things. They in some sense can happen. The thought can even efficiently cause the happening. Still, the happening cannot relate to the thought such that I can act in this way. A principle that governs our action cannot require such things. It instead requires something of me only if I can act from my understanding of the action. I must be able to act from thought about the action. Call this notion of possibility 'practical cognitive possibility'. It is a bit hard to state precisely, as I shall explain at the end of the next section. For now, though, say that the condition uses a notion of possibility that is at least as restrictive as practical cognitive possibility relative to the kind of being subject to this principle.

As the name indicates, this notion provides the correct interpretation of the modal vocabulary in the condition. It in fact explains why the less restrictive notions of possibility condition principles that govern my action. Just think about being subject to a principle that requires an action that is impossible in any of those ways. How could I act from thought? Some are unthinkable. No acting from thought where there is no thought, and I thereby cannot comply with a principle that requires an unthinkable action of me. Some are thinkable but not doable by me because they are beyond human abilities. No acting from thought where there can be no act, and I thereby cannot comply with a principle that requires an undoable action of me. Others are thinkable and can happen but not in a such way that the happening is a doing of mine. No acting from thought there either. In none of these cases can I self-consciously comply with a principle that requires such an action of me. I thereby cannot be subject to

that principle. If I could, normative principles for my action would be foreign to me. They would be insensitive to the nature of human action, the nature of what they govern. I could not base my action on them. The correct interpretation of the condition thus uses the notion of practical cognitive possibility relative to the kind of being subject to the principle. It captures a minimal sense in which normative principles for my action are internal to me.

3. PROBLEMS WITH POSSIBILITY

Although an argument from cases might clarify the proper interpretation of the modal terminology in the practical cognition condition, it does not explain why that interpretation is correct. Moreover, this argument has two problems. For one thing, the same considerations might seem to license more restrictive notions of possibility that conflict with the objectivity of normative standards for our action. For another, the kind-relative notions of physical and practical cognitive possibility are too permissive, at least without further explanation. Let me explain these problems in turn.

Start with a notion of ‘motivational possibility’ that says that I can act in some way only if so acting is compatible with my motivational tendencies. By ‘motivational tendencies’ I mean the stuff that I tend to factor into decision-making, allowing for means-end reasoning and calculative thought about ends themselves from the standpoint of other ends.²⁰ Let this notion also include propositions about

²⁰ Bernard Williams defends a similar interpretation. He starts from the premise that “if there are reasons for action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action” (Williams [1980] 102). While this condition might seem different from the practical cognition condition, that appearance disappears once what it is for us to act for a reason is explicit. Because we act from thought, a capacity to act for a reason is a capacity to act from my grasp of that reason—a notational variant on acting from my understanding of what a principle requires, permits, and prohibits. This premise is distinct from his internalist conclusion, which follows only given certain views of action and agency. I cannot get into these issues here. For present purposes, it is enough that he denies that requirements govern our action independently of our idiosyncratic psychologies. According to him, the psychopath that I am about to discuss does not act as she ought not to act. In this way, his commitment to the internality of normative principles for our action leads him to deny their objectivity. I shall argue that we need not abandon one in favor of the other.

how I never consider certain stuff. Now think about morally and prudentially insensitive psychopaths. They do not understand why moral and prudential considerations might bear on the question ‘What am I to do?’ They do not regret not acting on those considerations or think of it as a failure. They do not alter their behavior in response to the way that others react to immoral and imprudent conduct. Motivational possibility thereby rules out this person’s acting on moral or prudential grounds. An interpretation of the practical cognition condition that uses this notion of possibility precludes any principle that includes moral or prudential requirements from governing our action.

The considerations about acting from thought might seem to support this interpretation of the condition. In a recognizable sense, a psychopath cannot act from her sense of what a principle requires, permits, or prohibits if it requires, permits, or prohibits moral or prudential actions. Even if she could understand what it *would* require, permit, and prohibit, she cannot understand *that* it requires, permits, and prohibits her to act in various ways. She cannot understand its authority for her. How can she act from thought in these ways but not in the ways that less restrictive notions of possibility exclude? It might then seem like the notion of possibility in the condition must be at least as restrictive as motivational possibility relative to the agent in question.

However, this interpretation lets me out of normative requirements too easily. If I am not moved by knowledge of your good, I cannot be subject to a principle that requires me to act for it or to constrain my action in light of it. I am not subject to moral requirements, and I do not act badly—imperfectly, defectively, or deviantly—when I act immorally. If I am not moved by knowledge of my own good, I cannot be subject to a principle that requires me to act for it or to constrain my action in light of it. I am not subject to prudential requirements, and I do not act badly when I act imprudently. If I am not moved

to revise intentions when they conflict, stick to a judgment in the face of temptation, or act from practical judgment, I cannot be subject to a normative principle that requires these things. I am not subject to rational requirements, and I do not act badly when I am akratic, incontinent, weak-willed, or otherwise irrational. If motivational possibility provides the proper interpretation of the modal vocabulary in the practical cognition condition, I do not fail to act as I ought to act in any of these cases.

Psychopathy, though, is pathological if it is anything, as is any form of immorality, imprudence, and irrationality. Like psychopaths, lions, tigers, and bears cannot grasp moral, prudential, or rational requirements and hence cannot act from such grounds of action. Unlike psychopaths, this is not a defect in their development or exercises of their capacity to act. They are not members of a moral, prudential, and rational kind of being. Psychopaths are not a distinct kind of animal. They are human beings with defective or degraded capacities. They fall under a normative standard with us, though they consistently deviate from it. If motivational possibility conditions normative principles for our action, though, a psychopath does not exercise her capacity to act defectively. Her capacity to act is not defectively developed. She does not fall under a normative standard with us.

Of course, motivational possibility is not maximally restrictive. Each more restrictive notion determines an interpretation of the condition in terms of a notion of what 'I can do', all the way to a full description of the world at the moment of action. The idea of acting from understanding might seem to support each subsequent interpretation, to the point that it might seem that the only thing that I *can* do is what I *actually* do. The temptation of internality thereby seems to erode the basis of the objectivity of the basic normative standard for our action. The condition seems too robust to condition normative principles for our action.

Hold onto objectivity instead of internality, though, and we seem to swing to the view that normative principles for my action are insensitive to my abilities. Say we deny that these too restrictive notions condition normative principles for our action. If I am required to act in some way for some purpose right now even though so acting is motivationally impossible for me, why think that normative principles for my action are sensitive to my abilities at all? After all, the psychopath no more understands a principle that requires motivationally impossible actions than I understand a principle that requires practically cognitively impossible actions and all the rest. Denying that motivational possibility conditions normative principles for our action seems to imply that weaker notions also do not. The allure of objectivity thereby seems to corrode the basis of internality. The condition seems too empty to condition normative principles for our action.

What supports a certain notion of possibility thus seems to support too much. What undermines other notions seems to undermine too much. What explains which interpretation of the condition is correct? What shows that normative principles for our action can be internal objective standards for us?

I shall address this issue later. Let me turn to the problem about physical and practical cognitive possibility relative to the kind of being subject to the principle. The simplest interpretations of these phrases are too permissive to feature in the condition. Take physical possibility. I can be blown to bits by the bomb, be tossed and thrown by a tornado, and be mangled and mashed by a meteor. Not so were I the number 3 or the concept HORSE. None of these things, though, are physically possible in the sense relevant to the condition. A principle that governs my action cannot require them. I can move my arms, swivel my head, and all that jazz. Not so were I the number 3 or the concept HORSE. These things are physically possible in the sense relevant to the condition. A principle that governs my action can require

them. Hence, only some of what is possible given my physical nature is relevant to the correct interpretation of the condition. But what unifies what is physically possible for me in the relevant sense and distinguishes it from other things that are possible given my physical nature?

Likewise for practical cognitive possibility. I can stumble and tumble down the alley because I am distracted by the neon flashing lights of cosmic bowling. I can mistakenly or accidentally shoot your donkey when I mean to shoot my own. I can inadvertently knock over the candle while hand-waving my way through a paper presentation. Although these movements depend on thought, they are not relevant to the condition. A principle that governs my action cannot require them. In contrast, I can walk to the store because I know that my partner needs manchego for the risotto. I can jog in the park because I know the importance of exercise for health. These movements that depend on thought fall under the relevant notion of practical cognitive possibility. A principle that governs my action can require them of me. But what unifies what is practically cognitively possible for me in the relevant sense and distinguishes it from other bodily movements that depend on thought?

4. CAPACITIES AND THE STRUCTURE OF POSSIBILITY

I will return to those problems in the next section. In this one, I lay out the metaphysics of capacities and the structure of possibility that comes with it. As with the rest of the paper, I cannot consider questions and objections. I shall just present the aspects of this metaphysics needed in order to explain the conditions of agency and the theses about our agency.

A capacity is a potentiality, but not all potentialities are capacities. A mark of a capacity is that its actualizations divide into the good and bad, the successful and unsuccessful, the perfect and defective,

the complete and incomplete. The actualization is something that its bearer does, not something that happens to her. Only some things can be active in the relevant sense, and nothing is active with respect to everything that can happen in its history. For example, I have the capacity to act but not the capacity to be blown up by the bomb even though I am potentially blown to bits by the bomb, and I have the capacity to breathe but not the capacity to suffocate even though I am potentially suffocated. The primary use of 'capacity' picks out the capacities of living beings, with a derivative use picking out the capacities of artifacts. The concept lacks application with respect to non-living non-artifactual stuff. It is inactive, and the manifestation of potentialities thereby do not divide into the good and bad, the perfect and imperfect, the complete and incomplete. They just are.

'Capacity' thus encompasses what a long tradition calls active and passive powers, where an active power gives form to its object while a passive one takes form from its object. My capacities to hear and see are passive powers, my capacities to digest and act active. Both kinds of powers differ from my disposition to dissolve in lye and my disposition to fall to the Earth at the rate of the gravitational constant. In my terminology, 'potentiality' picks out the genus of which 'capacity' and 'disposition' are species, and 'capacity' picks out the genus of which 'active power' and 'passive power' are species.

A capacity allows for a manifold of particular bearers. A few billion human beings have the same capacity to act, with no upper bound. It is not like the stock will run dry and some must go without. Though resources are finite and can only support so many of us, the capacity itself does not limit the number of its bearers in the way that the number of seats in Senate limits the number of senators. Likewise, we can exercise the same capacity indefinitely many times. I exercise my capacity to act many times each day, with no in principle upper bound. I cannot exhaust the stock of exercises and be left

running on empty in the way that I can exhaust the number of attempts to enter my password and be left locked out of my account. Although we can only do so much in a day, in a life, the capacity itself does not limit the number of its exercises. A capacity allows for a manifold of exercises by a manifold of bearers.

A capacity is such that a single principle describes its nature and is thereby normative for its development and exercises. 'Capacity' thus picks out a type of kind whose members establish normative standards for their particulars. This type of relationship does not hold with respect to all kinds. While the nature of squares establishes standards for whether something is a square, it does not establish standards for whether something is a good or bad square. Likewise for potentialities that are not capacities. The idea of an internal normative standard does not make sense here. Not so with respect to capacities.

A capacity is a potentiality *to do* something, and its principle distinguishes it from others by describing an activity. My capacity to walk is different from my capacity to jump given how the way that we walk differs from the way that we jump. Those differences thereby distinguish the standard for developing and exercising the former from the standard for the latter. Each of them, though, is a more determinate version of my capacity to act. Walking and jumping are particular ways of acting.

To understand this determinable/determinate relationship, think about my capacity to speak. I am born with it. I develop it by learning to speak at least one language, and I exercise it by speaking a language. The only way to develop and exercise the determinable capacity is to develop and exercise at least one of its determinates. One way to think about this stuff is to treat 'German', 'English', and the others as adverbial modifications of 'speak' in 'my capacity to speak'. In the same way, 'slowly' and 'quickly' can modify 'walk' in 'my capacity to walk'. My capacity to speak, walk, jump, and others are likewise determinate versions of my capacity to act. After all, I can only develop and exercise my capacity

to act by developing and exercising capacities to act in various ways. As with my capacity to speak, one way to represent this determinable/determinate relationship is to treat ‘walking’, ‘jumping’, ‘speaking’, and the like as adverbial modifications of ‘act’, as in ‘my capacity to act by walking’.

Representing the relationship between my capacity to act *as such* and my capacity to act *in some way* as a determinable/determinate relationship is not mere scholasticism. The determinable restricts its determinates. Take my capacity to walk forward, itself a determinate of my capacity to walk, itself a determinate of my capacity to act. If to walk forward is in part to put one foot in front of the other, the principle of any determinate of it must include this aspect. This principle can modify the first one in various ways—‘put one foot in front of the other at a leisurely pace’ is a different determinate than ‘put one foot in front of the other at a cautious pace’. But it must stay true to the determinable. This is why strutting, slinking, and sauntering are determinates of my capacity to walk but walking while remaining still or walking forward backward are not.

Once again, one way to represent this relationship is to treat the descriptions of the determinate activities as adverbial modifications of the determinable. Only certain adverbs attach to the infinitive *to walk* to yield a semantically well-formed verb phrase. ‘Slowly’, ‘quickly’, ‘forward’, ‘backward’, ‘quickly forward’, ‘quickly backward’, ‘slowly forward’, and ‘slowly backward’ do. ‘Quickly slowly’, ‘forward backward’, ‘while remaining still’, and ‘in a zero gravity environment without mechanical aid’ do not. These phrases try to describe an activity that is internally contradictory in one way or another and hence is not a real activity, or at least not a human activity. We cannot have a capacity to act in these ways. To set off so to act is to automatically fail in the way that to set off to square the circle using only compass and straightedge is to automatically fail. Just try to make a go of it. I would then be exercising my

determinable capacity to act in a way doomed to failure because I in principle lack the determinate capacity to act in that way. Given the nature of a capacity, then, only certain determinates of it are possible.

The principle of our capacity to act likewise restricts its determinates. If this principle includes moral, prudential, and rational content, determinates of it obey moral, prudential, and rational requirements. A capacity to act immorally, imprudently, or irrationally would be as impossible for us as a capacity to walk forward backward. The phrases 'act immorally', 'act imprudently', and 'act irrationally' would try to describe activities that are internally contradictory in one way or another and hence are not real activities, or at least not human ones. We cannot have a capacity to act in these ways. To set off to so act is to automatically fail in the way that to set off to sail to the edge of the earth is to automatically fail. Again, it is not that I am not exercising my determinable capacity to act. It is instead that I am exercising it in such a way that success is impossible because I in principle lack the capacity to act in this way. Just as I can be on the boat sailing while not possibly succeeding in sailing off the edge of the earth, so I can be disregarding necessary means and undermining my good and yours as well while not possibly succeeding in acting irrationally, imprudently, or immorally. While I shall not here argue that the principle of our capacity to act as such in fact establishes these requirements, this metaphysics of capacities explains the possibility of such requirements on our action as such.

Two sorts of possibility come with any capacity. Human beings, like all of the living, are not passive in the face of the world. Of all the happenings in my life, some are my doings. A capacity is a potentiality to do something. To exercise it is to be active in the broad sense that includes seeing, hearing, digesting, and acting. Take my capacity to walk. Because I have developed this capacity, I can

walk. I am in this way differ from other organisms. No being without a capacity to walk can walk. No being without a human capacity to walk can walk as we do. In this way, a capacity makes possible its exercises, which are subject to normative standards that its nature establishes. It also makes possible characteristic defects or imperfections in its exercises. If I can walk, I can trip, slip, stumble, or tumble. Any being who can walk can err in these ways, and only a being who can walk or act in a similar way can err in these ways. Walking is doing what I have a capacity to do. Tripping and the like are deviations or imperfections in walking. They constitute failures of those exercises if they are severe enough. A walk to the store in which I get to the store by walking is successful though imperfect if it includes a stumble and other imperfections along the way. A walk that ends with me flat on my face with a fractured ego and an even worse leg is a failed exercise of that capacity.

Whereas walking is something that I can *do* because I have a capacity to walk, stumbling is something that *can happen* because I have that capacity. My capacity to walk thereby brings with it two sub-classes of what *is possible* with respect to it, what *I can do* because I have it and what *can happen* because I have it. This terminology is a bit uncomfortable for two reasons. First, active and passive powers are capacities, and yet I am in general the patient, not the agent, of a causal relation when I exercise a passive power. This explains the awkwardness of saying that hearing, say, is something that I do or is an activity. Second, 'what can happen' might sound like it only picks out things that come from outside the exercise of the capacity, not things that can be done deliberately. Both of these impressions are misleading. 'What *I can do* because I have a capacity' picks out the aspects of the exercises that realize the nature of the capacity and contribute to a successful exercise of it. It picks out the way that things can go well with respect to the capacity. 'What *can happen* because I have a capacity' picks out the aspects of

exercises that interfere with realizing the nature of the capacity and contribute to a failed exercise of it. It picks out the way that thing can go badly with respect to the capacity. What I can do because I have a capacity are the perfections, what can happen the imperfections, of an exercise of that capacity. Imperfections can come from outside the exercise itself as when a tornado ends my walk to the store. They also can be internal aspects of the exercise, as it were, in the way that the imperfection of swimming to the bottom of a volcano comes from inside the exercise.

Although these sub-classes of what is possible with respect to a capacity are a package deal, they are not on a par with each other or with respect to their relationship to the capacity. A capacity is a potentiality to do something. Its nature depends on what it is a potentiality to do. What differentiates my capacity to walk from my capacity to sing is, after all, the difference between walking and singing. Because of the nature of walking, tripping and the like are possible imperfections in exercises of my capacity to walk. Because of the nature of singing, my voice cracking and not holding the tune are possible imperfections in exercises of my capacity to sing. If any of these imperfections are severe enough, I fail in those exercises. The perfections of exercises of the capacity are in this way prior to the characteristic imperfections in and failures of those exercises. Walking is an actualization of my capacity to walk, whereas tripping is an imperfection or defect in its exercise that is possible because of the nature of walking. Tripping is a defect in walking. It is in this sense derivative of walking and understood in terms of walking. A capacity to walk is, after all, a capacity to walk, not a capacity to walk and also to trip, be run over by a bus, and every other way for its exercise to be imperfect. A capacity thus distinguishes what I can do because I have it from what can happen because I have it. What can happen depends on what I can do because what can happen are interferences with, interruptions in, and other imperfections

of what I can do.²¹

Many capacities do not come fully formed. I am not at birth able to do everything that I have a capacity to do. I am not able to speak, walk, or act at birth even though I have these capacities. I need to develop them if I am to exercise them, through growth, education, training, reflection, and everything else that goes into a life. In order to account for this aspect of capacities, the metaphysics must include the state of development of the capacity in addition to the capacity and its exercises. To put the same point a different way, this metaphysics must distinguish two ways of *having* a capacity. In one sense, I have a capacity because I am a being of a certain kind. I have a capacity to walk because I am a human being.²² If I were a member of another species, I would have a different capacity to walk or none at all. I have a capacity in a different sense when I have developed it such that I am able to exercise it in at least some of the ways appropriate to it. It is the same capacity at a specific stage in its development. This tripartite distinction between the capacity, its state of development, and its exercises is crucial for understanding the conditions of agency. Let me explain it in more detail.

Consider the baby. A human infant has the capacity to walk but not the capacity to fly. Everyone recognizes this distinction, or at least everyone recognizes a difference between her relationships to

21. In addition to errors in the exercise of a capacity, I can also fail to exercise a capacity. The metaphysics of capacities thus must explain how a capacity brings with it standards for when to exercise it in addition to standards for how to exercise it. However, I ignore this aspect of capacities because it does not matter for the conditions and theses of interest.

22. In a sense, every capacity that I can possess comes with being a member of a species since they are all determinates of my basic capacities. In another sense, only certain capacities come with being a member of a species. I by nature have a capacity to act and the more determinate capacity to walk. I do not by nature have the capacity to play video games even though I can train into that activity given my basic capacities. I do not have an account of which determinates come with our nature and which do not. The important question is whether failing to develop the determinate constitutes a deprivation or defect. A rough test is to think about whether the activity in question is something that any of us must be able to do in order to live a human life well. This test is not foolproof. It relies on a conception of our kind of life as present in the many ways we live that distinguishes the essential and non-essential aspects of those ways of life. I do not have a procedure to recommend for answering these questions other than 'think hard and carefully'.

walking and flying. Of course, she cannot yet walk. If you ignore the 'yet', her relationships to walking and flying are similar. That 'yet', though, is all important. Although she cannot yet walk, she will be able to walk if the normal course of development takes place. She will be in that position because she is a human being or, in other words, because of her nature. This capacity by nature develops in her over time. Not so with flying. You need wings or something like them for that.

Any view of capacities needs these three elements. To see why, try to get by with two. If a single principle describes the nature of a capacity and is thereby normative for its exercises, what principle describes the baby's capacity to walk? If it describes the current state of that capacity, it hardly is a capacity to walk, and she does not relate to walking and flying differently. After all, she can neither walk nor fly right now. Moreover, if a principle describes a capacity in its current state, every capacity that I possess meets its principle. This has three problems.

First, the ideas of a capacity developing and degrading become unintelligible. To develop a capacity is to come to be able to exercise it in ways that by its nature I should but previously I could not. For it to degrade is for it to go in the other direction. Development and degradation progress along an intrinsic spectrum of perfection. If the principle describes the capacity as I have developed it, though, there is nowhere to go. Change might happen, but it would be change from one capacity to another, not development or degradation of a capacity. Second, if the principle describes my capacity in its current state, imperfections in my development will condition the normative standard for its exercises. My deficiencies and limits become normative standards for my action because nothing determines whether my development is incomplete. Finally, if the principle describes the current state of the capacity in the infant, she and I cannot possess the same capacity. After all, I can walk. Likewise, my capacity to walk will

not be the same as yours. Perhaps the *content* of the descriptions of our capacities will be the same, but that content derives from different things. The principle does not describe something present in each of us. Or, to put it another way, we are not distinct bearers of the same capacity, and a capacity does not allow for a manifold of bearers. It only allows for one.

Still, if there are only two parts to the explanatory nexus, what happens to our differences with respect to this capacity? If the principle describes the capacity to act as we possess it because we are human, the infant and I have the same capacity. She and I also might not exercise it in the same way in the same circumstances. We might sit there splayed out on the floor looking at you all goofy and disinterested while you plead and try all manner of persuasion to get by us and continue down the toy aisle. Only one of us, though, is failing to exercise our capacity to act or exercising it badly. Only one of us has developed this capacity such that we are able to exercise it. Hence, there must be three elements of the explanatory structure of capacities, the capacity as I possess it because I am a human, the capacity as I have developed it, and its exercises. I possess a capacity because of my nature. I develop it over time in an upbringing, and along this spectrum comes the possibility of exercising it in various ways to various degrees in various circumstances. And I exercise it in various moments throughout my life.

To sum up, the principle of our capacity to act describes that capacity as we possess it because we are human beings and thereby establishes a normative standard for its development and exercises. It sets the standard according to which change is development or degradation and according to which exercises are perfect or imperfect and success or failure. To exercise it well is to act as this principle requires or permits and refrain from acting as it prohibits from my sense of what it requires, permits, and prohibits. To exercise it imperfectly is miss the mark along some dimension of this exercise. To fail is to

not pull it off at all. This capacity thus structures two classes of what *is possible* with respect to the capacity, the class of what *I can do* because I have it and the class of what *can happen* because I have it. Although these possibilities come with each other, the former subclass is prior to the latter because what can happen are per se imperfections in what I can do. They are per se interferences or interruptions in the exercise of a capacity. This is why I have a capacity to walk, not a capacity to trip, even though the possibility of tripping comes with the possibility of walking. It is why, assuming that we are moral, prudential, and rational beings, we have the capacity to act morally, prudentially, and rationally, not the capacity to act immorally, imprudently, and irrationally. Morality, prudence, and rationality mark perfections of exercises of our capacity to act, immorality, imprudence, and irrationality imperfections that constitute failures in exercises if they are severe enough.

5. WHAT CAN I DO?

In the rest of this paper, I use this metaphysics of capacities in order to explain the correct interpretations of the conditions and theses of interest. I begin with the practical cognition condition because the notion of possibility in it is explanatorily prior to the notions in the error and alternate possibility conditions.

Start with the relationship between these conditions:

The Success Condition A principle governs my action only if I *can* comply with it.

The Practical Cognition Condition A principle governs my action only if I *can* self-consciously comply with it—that is, only if I *can* act as it requires or permits and refrain from acting as it prohibits from my sense of what it requires, permits, and forbids.

The practical cognition condition partially explicates the nature of compliance in terms of *acting* as the principle requires or permits and *refraining* from acting as the principle prohibits. These ideas pick out

things that I *do*, not interferences or interruptions in doing something. They pick out possible intentional actions of mine. This condition thus uses the notion of what I can do because I have a self-conscious capacity to act, not the notions of what can happen or what is possible because I have it. Since to have a self-conscious capacity to act is to have a capacity to act from thought about action, what I can do because I have this capacity is what is practically cognitively possible for me. And this metaphysics specifies this notion in a way that meets the two challenges from earlier. Let me explain.

One challenge is to show that physical and practical cognitive possibility do not include everything possible given my physical nature and every movement that can depend on thought. Otherwise, I can be required to be mashed by a meteor or to trip because I am distracted by neon lights. They are things that can happen to me given my physical nature and my ability to act from thought. The proper notions of ‘physical possibility’ and ‘practical cognitive possibility’, though, use the notion of what I can do because I have a self-conscious capacity to act. Things that can happen because I have it thereby cannot govern my capacity to act. They are not physically or practically cognitively possible in the relevant sense. And this restriction is explained by the nature of our capacity to act. After all, the question is about our actions, which are things we do, not possible imperfections in actions. Because we act from thought, I can be required or permitted to act in some way only if it is presented under a description such that intentionally acting that way is possible for me. Likewise, I can be prohibited from acting in some way only if it is presented under a description such that intentionally refraining from acting in that way is possible for me. Only if a principle respects these limits can it govern *our action*.

The other challenge is to show that this interpretation shows that the basic normative standard for our action can be an internal objective standard. The key here, I shall argue, is the distinction

between my capacity to act as I possess it because I am a human being and that capacity as I have developed it. I will argue that the practical cognition condition uses the notion of what I can do given my capacity to act as I possess it because I am a human being. It thereby vindicates the internality and objectivity of the basic normative standard for our action. Later in the paper, I shall argue that the alternate possibility condition uses a notion of possibility relativized to the state of development of my capacity to act.

Because my capacity to act is self-conscious, its principle must be appreciable from the first-person perspective. I must be able to act from my understanding of it. Yet I am not always in a position to understand it. I might not now be in such a position. 'Appreciable', after all, is a modal term. A fragile vase is not always in a position to break. It might not now be in such a position if it is under protective wrap with an automated defense system and a security detail. Still, it is fragile because fragility is an intrinsic property of the vase, not an extrinsic property that depends on its circumstances. The metaphysics of capacities provides the model for understanding how the principle of our capacity to act is likewise appreciable from the first-person perspective even if I cannot now appreciate it.

I possess a self-conscious capacity to act because I am a human being. I have possessed it since birth even though I could not then exercise it, let alone guide myself in those exercises by its principle. I have developed it, becoming able to exercise it in various ways. This progression includes developing my physical abilities. At one point, I could not move my fingers in the ways needed to tie my shoes. At one point, I could not track flying balls and orient myself in order to catch them. I picked up these skills early and have refined them. Not perfectly, mind you, as the departmental softball team discovered to their regret. This progression also includes coming to understand grounds for acting. For instance, while

humans quickly understand hunger as a ground of action, it takes time to learn when relieving hunger is improper. It takes longer still to understand health, beauty, knowledge, justice, honesty, and other determinate principles of prudence and morality. It takes a lifetime to understand rationality, prudence, and morality themselves as proper principles of choice for actions. Even then, it rarely works perfectly. This stuff and more goes into developing our capacity to act.

Part of developing a self-conscious capacity is developing the ability to reflect on it and on my exercises of it. Through reflection, I can come to understand the principle of this capacity in a way that lets me act from my understanding of what it requires, permits, and prohibits. In understanding this capacity, I understand myself as one of a manifold of beings who by nature possess this capacity and so by nature fall under a single normative order. In doing so, I understand that this capacity develops over time. I thereby understand that at any time I might not recognize the proper ends of action or the proper principles of choice for actions. Such ignorance is a characteristic imperfection of the development of this capacity that can lead to failures to develop it and exercise it when and as I should.

Because we have the same capacity to act, we fall under a shared normative order. Because this capacity by nature develops over time, we might be at different points along this developmental spectrum. To some extent, my state of development determines which of the requirements that follow from the basic normative principles for our action apply to me. As a baby, I do not fail to exercise my capacity to act when I do not recognize your claim to a clear path in the toy aisle. As a reasonably well-developed adult, I do. My state of development itself, though, is also an object of normative assessment. My development is incomplete if the proper description of my capacity as it is developed is not identical to the proper description of my capacity to act as I possess it because I am human. My development is

defective if it is incomplete and I should have developed it further—if it is less developed than is proper at this stage of life. This idea is rough since so much that goes into developing our capacity to act depends on social and political circumstances whose influence is hard to notice, quantify, control, and counteract.

These ideas explain why the psychopath acts as she ought not to act in a way that the elephant who causes just as much pain and torment does not. The psychopath has a defectively developed self-conscious capacity to act. Because of something pathological in her biochemistry or upbringing, she is ignorant of grounds for action and principles of choice that she by nature should understand. When she acts against her own good or my good, she acts against what the principle that governs exercises of her capacities to act requires of her. Because of her pathology, she is not now able to understand what this principle requires, permits, and prohibits of her and act from her sense of it. Yet this principle is appreciable from her first-person perspective in the relevant sense. For one thing, to the extent that she can understand herself as acting, she can understand that she is one of a manifold of beings who falls under a shared normative order. Even if she cannot grasp the details of that order, she can understand its possibility. If she has sufficiently developed the capacity, she might understand that her development can be incomplete and hence that she might not be able grasp some aspects of its principle. She might even understand the possibility of pathological defects that get in the way of developing it. She cannot get herself aright with respect to this principle. Still, she falls under it because of the nature of her capacity to act, which she shares with us even if she cannot come to much reflective understanding of it.

For another thing, it is of the nature of her capacity to act that she be able to appreciate this principle. If her development of it goes well, she will appreciate this principle, dimly at first and more

clearly over time. In this way, the psychopath, like all of us, is the ground of the normative standard that governs her action as such. Her capacity to act is the source of the principles that govern its exercises, and she will understand them if it develops as it by nature develops. The principle that governs her exercises of it are thereby not alien to her or forced upon her. She, like all of us, is not always in a position to recognize them and act accordingly. She is not now in such a position, and her pathology will keep her from getting there. These principles then *seem* to her alien, as moral and prudential principles often seem in our development and in trying situations. Yet the principles are not *in fact* alien to her or to us, not in a way that matters for their internality, because our capacity establishes the standards that govern its exercises and development. She has the same capacity as we have, with the same normative standard, because she is human. She is one of us.

The distinction between two ways of possessing a capacity thus provides an account of the internality of the basic normative principle for our action that does not make it subject to individual idiosyncrasies. It likewise shows that this principle can be objective for us. The nature of our capacity to act sets the standards for each of us, and our capacity is sensitive to our abilities because it is our general ability to act. Yet it is not necessarily sensitive to our idiosyncrasies. In particular, it is insensitive to deficiencies in our development of that capacity like, say, an inability to recognize moral and prudential grounds for action. Although sensitive to the nature of that capacity as you and I possess it because we are human, it is not necessarily sensitive to the capacity as we have developed it, to our decisions about whether to develop our capacities in certain ways, or to our exercises of it if they violate its principle. It sets the standards for development and exercise independent of our particular developments and exercises. The practical cognition condition thus uses the notion of practical cognitive possibility. It

thereby shows that the basic normative standard for my action can depend on my nature but not on my imperfections. Hence, the standard can be an internal objective normative standard for our action.²³

6. I CAN ERR

The Error Condition A principle governs my action only if I *can* deviate from it.

The first cause of recent reflection on this condition is Douglas Lavin's 'Practical Reason and The Possibility of Error'. He initially contrasts two interpretations, the *logical* and the *imperative*. The logical interpretation "says that an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle" (Lavin [2004] 426). The imperative interpretation says that "an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle and it is possible for the agent to do it" (ibid. 427). The logical interpretation says that any principle that governs my action allows for deviation. It does not say that each agent subject to the principle can deviate or even that any being actually subject to that principle can deviate. It only says that some possible agent subject to the principle can deviate. In contrast, the imperative interpretation says that each agent subject to a principle can deviate. Hence, whereas the logical interpretation says that a principle can govern my action only if some possible agent can deviate from it, the imperative interpretation says that a principle can govern my action only if I can deviate from it.

Lavin further distinguishes between the *weak* and *strong* imperative interpretations. The weak

23. I further consider whether this account captures the relevant sense of internality in §7. I there distinguish a question about whether I succeed or fail in an exercise of my capacity to act from a question about whether I am responsible for that success or failure. The notion of internality that this account does not capture goes with the latter question, which is the province of the alternate possibility condition, not the success or practical cognition conditions.

version says that “an agent is subject to a principle only if there is some kind of action such that if the agent did it she would violate the principle and it is possible for the agent to do it, and it doesn’t matter why the agent does it” (ibid. 436). The strong version says that “an agent is subject to a principle only if there is some kind of action such that if the agent did it she would violate the principle and it is possible for the agent to do it and her doing it would be a genuinely practical error” (ibid.). The contours of these interpretations come out in cases that satisfy the weak interpretation but not the strong one. Take a view that says that I can only make errors in action if they derive from errors in exercises of theoretical reason. That is to say that I cannot make basic errors in exercises of my capacity to act. Maybe I mistakenly think that someone who is playing a game in the lake is drowning. I decide to jump in order to save him and end up getting a cramp halfway there and putting myself in peril. I here err in exercising my capacity to act, both in the thinking and doing. Those errors, though, are not basic practical errors because they derive from an erroneous exercise of theoretical reason. The existence of a being who can *only* make derivative errors of this kind is compatible with the weak imperatival interpretation but not the strong one. In contrast, if a principle can govern my action only if I can commit basic practical errors with respect to it, the strong imperatival interpretation is true. Since I shall question what is needed for an error to be a basic practical error in a bit, I leave this interpretation in this schematic form for right now.

Any principle from which I can deviate by making a genuine practical error is a principle from which I can deviate by making an error of any kind. Any principle from which I can deviate by making an error of any kind is a principle from which some possible agent can deviate. The question is then not which interpretation, if any, is true. Lavin explicitly criticizes only the strong imperatival interpretation.²⁴

24. Lavin’s criticisms, if sound, also refute weak imperativalism, as Lavin recognizes. For one thing, he criticizes strong imperativalism for being incompatible with perfect rational agency. However, perfect rational agency is incompatible with

In effect, he argues that the strong imperatival interpretation transposes the error condition into the key of self-conscious agency in the way that the practical cognition condition transposes the success condition. This transposition turns the error condition into the following:

The Willful Violation Condition A principle governs my action only if I *can* self-consciously deviate from it—that is, only if I *can* refrain from acting as it requires and act as it prohibits from my understanding of what it requires and prohibits.²⁵

I agree with Lavin that this interpretation is bad. I shall argue, though, that the strong imperatival interpretation is identical to the willful violation condition only if the error condition uses the notion of what I can do because I have a self-conscious capacity to act. This interpretation is incorrect, though, because it treats error as perfection, not imperfection or failure, in my exercise of a capacity. The proper interpretation instead uses the notion of what can happen because I have a self-conscious capacity to act, which vindicates the strong imperatival interpretation by disentangling it from the willful violation condition.

Lavin argues against the willful violation condition by criticizing Christine M. Korsgaard's

any errors in reason, including the kind of "error in action that derives from some defect in theoretical reason" that he considers while discussing weak imperativalism (Lavin [2004] 435-6). For another thing, he claims that

behind the strong imperatival interpretation is the thought that action is not the upshot of theoretical reason operating with a special set of contents ... but, rather, the upshot of a rational capacity of an entirely different kind. When conjoined with the intuition that kinds of rational capacity are in some sense tightly connected with kinds of error, we come close to the strong imperatival interpretation. However, we can capture the heart of the intuition without making imperativalist commitments by appeal to the much weaker *practical defect constraint*: for every kind of practical thought there is a unique kind of practical defect connected with it. It is a further and nontrivial step to claim that each bearer of a capacity for thought of the relevant kind must potentially exhibit such a defect or really be able to mess up in the relevant way. (ibid. 437n.18)

To the extent that the practical defect constraint threatens the plausibility of the strong imperatival interpretation, it threatens the weak one as well. Both 'make imperatival commitments' by saying that each bearer of a capacity for practical thought must potentially exhibit a characteristic defect. They just disagree on whether this defect is derivative or basic.

25. I do not know what to make of resisting permissible actions. Since nothing in my argument turns on this idea, I ignore it.

invocation of it in ‘The Normativity of Instrumental Reason’.²⁶ Let me explain why he thinks that she invokes it. The error condition is at the heart of her essay. She claims that “imperatives are addressed to beings who may follow them or not. ... [I]t *must be possible* for a rational being ... to disobey, resist, or fail to follow [any normative] principle” (Korsgaard [1997] 48; my emphasis). Her prime example of a failure is “something ... interfering with ... reason” such as when I am “rendered inert by depression, or paralyzed by terror, or ... simply unable to face” acting some way “because the means are painful” (ibid. 49). Although she does not reflect on the nature of capacities, these things mark imperfections of an exercise of my capacity to act. Aside from the fact that she calls them interferences with reason, just look at the verbs: I am *rendered inert* or *paralyzed*, not deliberately standing still; I am *unable to face* something, not deliberately looking away. These things happen in an exercise of my capacity to act. They are characteristic imperfections that might constitute failures if they are severe enough. They are not imperfections of an exercise of this capacity.

However, Korsgaard treats claims about possibility as on a par with claims about capacities. For example, she treats the claim that I am subject to a principle only if I am “capable of resisting” it as equivalent to the claim that I “can be subject to normative principles [in action] only if [I] can resist them, because without that possibility they cannot function as guides” (ibid. 52n.39). This might not seem like a problem since in chit-chat we can use ‘capable’, ‘can’, and ‘is possible’ to express the same thing. However, using them interchangeably lends itself to treating complying with and deviating from a principle as on a par with respect to our capacity to act, which Lavin correctly criticizes. Let me explain.

26. I shall avoid arbitrating between Korsgaard and Lavin. To just state my opinion, his criticisms apply to the letter of her view in that essay. Internal tensions in her essay, though, derive from her failure to reflect on the nature of practical possibility. Resolving them in the way that best fits with the aims and spirit of the essay requires attributing to her the view that I here develop. Hence, his criticisms apply to the letter but not the spirit of her essay, though the spirit is, so to speak, rarefied.

Although human beings can comply with and deviate from principles, to posit a 'capacity to resist' alongside a 'capacity to comply' is bad. If they are distinct capacities, to deviate from principles is to do one thing, to comply another, and each establishes its own standards of correctness for its exercises. In one respect, these capacities relate to each other like our capacities to stand and sit do. Although exercising one rules out exercising the other at the same time, sitting and standing are distinct kinds of activities. From the outside, a successful exercise of one looks like an unsuccessful exercise of the other. It is not, though, but something else entirely. Just so, if complying with and deviating from principles are different activities, each establishes its standard of correctness for its exercises. From the outside, a successful exercise of one looks like an unsuccessful exercise of the other. It is not, though, but something else entirely.

In another respect, though, these capacities cannot relate to each other as the capacities to sit and stand do. Sitting and standing are distinct kinds of human actions, and the capacities to do them are different determinates of our capacity to act. Capacities to comply with and deviate from principles are not likewise different determinates of a single capacity. Far from determining something in a way that remains true to it, the principles of these capacities instead concern how we are to relate to any principle whatsoever, including the principle that describes our determinable capacity to act. A capacity to comply in effect says 'do what principles require and refrain from doing what they prohibit'. A capacity to deviate in effect says 'do what principles prohibit and refrain from doing what they require'. What one prescribes the other proscribes. A capacity to deviate, then, cannot be a determination of our capacity to act. It inverts the normative standards that follow from that capacity, and in this way is not true to it. Since a capacity to comply is the mirror of the capacity to deviate, it likewise is not a determination of our

capacity to act. They instead are the basic rational or self-conscious capacities that condition all others. So, at any rate, you must say if you posit a capacity to deviate alongside a capacity to comply.

If I possess these distinct capacities, though, no principle can govern my capacity to act. Just think about the principles that come with this capacity. If I am exercising my capacity to comply, the principle of my capacity to act has authority for me. If I am exercising my capacity to deviate, that principle lacks authority for me. After all, if I am exercising my capacity to deviate, I exercise it well if I do not comply with the principle of my capacity to act. To deviate successfully is to refrain from doing what the principle of my capacity to act requires and to do what it prohibits. Moreover, since the capacity to deviate is a self-conscious capacity, I can understand its principle and guide myself by it. I thereby exercise it well if I deviate from my understanding of what the other capacity requires and prohibits given my understanding of the principle of that other capacity. To attribute to us a capacity to deviate on the grounds that a normative principle requires that each agent subject to it can deviate from it is then to say that the willful violation condition is true.

If that interpretation is true, though, we cannot be subject to normative principles in action. It is incompatible with the objectivity of the basic principle of our capacity to act because whether I am to exercise my capacity to comply or my capacity to deviate is not subject to a normative standard. It is a kind of arbitrary choice. Just try to establish a principle that governs them. Whatever it is, the capacity to comply says to comply, the capacity to deviate the opposite. I thus cannot be correct in determining whether to exercise one rather than the other. In other words, treating them as distinct capacities entails that whether I am subject to a principle depends on my arbitrary choice that cannot be right or wrong. Hence whether the principle of my capacity to act or its reverse govern in a particular case is arbitrary,

which is incompatible with the objectivity of the basic standard for our action.

This point leads to another criticism of the willful violation condition. Lavin argues that a capacity to deviate from principles alongside a capacity to comply with them is incompatible with the “conception of perfect rationality ... as a state of will nonaccidentally issuing only in correct action, though only contingently possessed by its bearer” (Lavin [2004] 451). If capacities to deviate and comply come with self-conscious agency, “the capacity to act just is, in part, the capacity to follow ... principle[s] [I]t is also true that the capacity to act just is, in part, the capacity to violate ... principle[s] But if it is constitutive of agency both to follow and to violate a principle, then we can no longer derive an intelligible commitment simply to follow from the nature of agency itself” (ibid. 454). If our agency brings with it these capacities, there is no way to privilege one of them. There is then no conception of perfect rational agency. Such an agent has incompatible capacities that she cannot jointly exercise and that she cannot hierarchically order. Which of the two she exercises cannot be correct or incorrect. Moreover, she can only exercise these capacities by exercising them in an executive role with respect to another self-conscious capacity, and she can only exercise other capacities under the auspice of one of them. She then cannot be correct or incorrect in whether and how to exercise her capacity to act and its determinates. Arbitrariness in this way corrupts. Yet in losing the notion of correctness we lose the notion of perfect rational agency.

Another argument to this same conclusion focuses on how the impossibility of correctness undermines the notion of a state of development of my capacity to act. Some changes are developments of our capacity and others are degradations because the principle that describes its nature is normative for its development. A fully developed capacity to act just is a perfectly developed capacity to act in

which the agent does not exercise her capacity in a way that is imperfect because of any internal deficiencies in those exercises. Without the possibility of correctness and perfection, a self-conscious capacity can simply change, with no normative standard on this change. It cannot develop or degrade and thus cannot have a state of development. A perfectly developed capacity, though, just is one where the principle that describes the nature of the capacity also describes its state of development. The loss of the possibility of correctness within the exercise of the capacity undermines the possibility of developing it. The loss of the possibility of developing it undermines the possibility of a state of its development. The loss of the possibility of a state of its development undermines the possibility of a perfect state of its development. And that just is to undermine the possibility of perfect rational agency.

Finally, the idea of a capacity to deviate is internally contradictory. The capacities to deviate and comply lack domain restriction. They thereby range over themselves, with the capacity to comply establishing complying with itself as success and the capacity to deviate establishing deviating from itself as success. However, the capacity to deviate then says to deviate from deviating from other principles. That is, it says to do what a principle prohibits and refrain from doing what a principle requires, and it also says to deviate from that order. To deviate from that order, though, is to do what a principle requires and refrain from doing what it prohibits. Far from establishing deviance as success, then, the capacity to deviate establishes compliance as success. Of course, it also says to deviate from that order, re-establishing deviance as success, and onward up the self-referential ladder. I thereby cannot have a capacity to deviate. It is internally contradictory. To set off so to act is to automatically fail in the way that to set off to enumerate the elements of the set of all sets that are not members of themselves is to automatically fail. There cannot be a capacity to deviate, and the willful violation condition is false.

The metaphysics of capacities, though, distinguishes the strong imperativist interpretation from the willful violation condition. Not everything that comes with a capacity is a perfection of its exercises. Just as walking and tripping come as a package even though only walking describes something that I can do, complying and deviating come as a package even though only the former describes something that I can do. Compliance partially describes the nature of our capacity and hence constitutes a normative standard for its exercises. That is what the practical cognition condition establishes. Deviation is a characteristic imperfection or defect in exercises that constitutes a failure of them if it is severe enough. That is what the error condition establishes. Korsgaard is right that the possibility of deviating comes with self-conscious agency while Lavin is right that it is not on a par with complying.

If complying with a principle describes the nature of our capacity to act and deviating is a characteristic imperfection in its exercises, the willful violation condition does not follow from the strong imperativist interpretation. More accurately, the strong imperativist interpretation is identical to the willful violation condition only if the only way for me to commit a basic practical error is for deviating to be something that I do, not something that can happen in exercises of my capacity to act. On a different understanding of basic practical error, the interpretations diverge. The way to go is to focus on mundane cases of error. Korsgaard and Lavin focus on cases of being unable to face taking the means to my end, akrasia, weakness of will, and diabolical evil. These kinds of errors, except maybe the last, are kinds of practical irrationality. It might seem like the debate about the strong imperativist interpretation is about whether I can be subject to a principle only if I can be irrational with respect to it. Lavin treats the debate this way, with justice given Korsgaard's language and examples. Some basic errors in exercises of our capacity to act, though, are not instances of irrationality. They exemplify the nature of error in a clearer

way than cases of practical irrationality. They should guide our understanding of the nature of practical irrationality.

When I am on the ground with a fractured leg and ego, I fail in my exercise of my capacity to walk. This failure might derive from a failure in theoretical reason—maybe I miscalculated the terrain. It might not, though. I might just trip. It might be a failure in the exercise of my skill of walking that is essential to this exercise of our capacity to act being an exercise of this determinate rather than another. Or maybe “the order is ‘Right turn’ and I turn left: no doubt the sergeant will insinuate that my attention was distracted, or that I cannot distinguish my right from my left—but it was not and I can, this was a simple, pure mistake. As often happens” (Austin [1956_a] 148n.1). In this way, it is a distinctively practical failure—a failure in the exercise of my capacity to act that does not derive from a failure in the exercise of another capacity. It is a distinctively practical failure that is not ‘irrational’. Moreover, it is a kind of basic practical error that comes with having a capacity to walk or a capacity to turn right. It is the kind of thing that can happen in exercises of those capacities even though it is not something that I do. Yet it includes nothing so fancy as to attract the gaze of practical philosophers.

The world of error includes mistake, accident, inadvertence, incompetence, inattention, neglect, clumsiness, carelessness, and much else besides. Not all need be instances of practical irrationality or diabolical evil. Of course, irrationality and evil fascinate and frustrate in a way that tumbling and stumbling never will. Their possibility comes with the possibility of rationality and goodness since what I can do brings with it the kinds of imperfections that can happen in exercises of that capacity. Yet the model on which to understand them is the same as the model on which to understand the other ways for me to err. Lavin does not disagree. He claims that a “capacity, ability, or power is ... always something,

in some sense, good. Indeed, ... the idea of a capacity or power to resist reason is a confusion on the order of the idea of a capacity not to see, or the treatment of blindness as itself a capacity” (ibid. 450n.48). Just as a susceptibility to blindness, whether momentary, temporary, or permanent, comes with the capacity to see, so a susceptibility to irrationality, evil, and the boring kinds of error comes with a self-conscious capacity to act. Just as blindness or failing to see something in front of you is not a *derivative* defect of our capacity to see, so error is not (always) a derivative defect in an exercise of our capacity to act. Still, blindness is not on a par with seeing. Likewise, deviating from principles is not on a par with complying with them, nor error generally with success. The metaphysics of capacities explains why. Complying marks what I can do. Deviating marks what can happen in exercises of my capacity to act. Deviance stands to compliance as imperfection to perfection.

Move to the possibility of perfect rationality. Lavin claims that imperativism cannot account for the possibility of perfect rational agency. This criticism does not turn on the distinction between basic and derivative practical errors. It applies to both imperativist interpretations. It in effect says that ‘the conception of perfect rationality as a state of will nonaccidentally issuing only in correct action, though only contingently possessed by its bearer’ is coherent only if a being can be subject to a principle even though she is in principle unable to deviate from it. That is to say that it in effect says that perfect rationality is possible only if the logical interpretation of the error condition is the strongest true interpretation.

However, the metaphysics of capacities explains this conception of perfect rationality while vindicating the strong imperativist interpretation. Just distinguish my capacity to act as I possess it because I am a human being and my state of development of it. Since we have a self-conscious capacity

to act, to exercise it perfectly is to act from my understanding of what I am to do. Although I possess this capacity by nature, I do not possess it in a developed state by nature. It by nature develops over time, with the perfect state functioning as the normative standard and aim of this development. If I fully develop it, I always know what I am to do and pull it off, at least absent interference by nature or others. I thus non-accidentally act correctly despite possessing this state of development contingently. After all, it is a developed capacity. Full development is, of course, long off. No one is likely to achieve it, at least given known social and political arrangements. Still, its possibility follows from the metaphysics of capacities even if possessing a self-conscious capacity to act entails the possibility of basic practical error.

Of course, this “terrestrial ... conception of a perfectly rational agent” is not everything everyone means by perfect rational agency (ibid. 450). The agent is not infallible, and in a few senses. First, I will likely err in exercises of my capacity to act while I am developing it. If I do not, it will be by luck or accident. Second, even if I fully develop my capacity, it will degrade from this state over time, at least absent medical and technological breakthroughs that will likely bring about the robot apocalypse. I will likely err as it does. If I do not, it will be by luck or accident. Third, even if my capacity is fully developed, I might fail in some exercises because of forces outside of my control. If I do not, it will be by luck or accident. A perfectly rational human being is, after all, a human being. I do not control everything, nor is everything that can affect the success of my action within my epistemic horizon. Finally, this conception of perfect rational agency is far from a conception of divine rational agency.

But, honestly, so what? An account of our agency and its normative standards need not concern itself with the possibility of a being whose capacities do not develop or degrade and who never requires cooperation from the world or anyone else. She would possess no unrealized potential, with no limits on

her abilities. It is thereby unclear what it means to say that she has a capacity, ability, or any other kind of potentiality. She is instead pure actuality, whatever that might mean. Far from being an essential component of a constitutivist metaphysics, “such a belief is not in line with the traditional beliefs enshrined in the word *can*: according to *them*, a human ability or power or capacity is inherently liable not to produce success, on occasion, and that for no reason (or are bad luck and bad form sometimes reasons?)” (Austin [1956_b] 166n.1). Sometimes I just mess up or get messed with. I am finite. My capacities thereby come with the possibility of success and error, as the ‘can’ of a capacity enshrines. As of human capacities, so of capacities generally. They bring with them the possibility of error, of internal and external interference, because they are potentialities. Non-accidental correctness in the exercise of them does not require the impossibility of error. It only requires that error does not come from an internal defect in the capacity if fully developed. A fully developed capacity is immune to internal imperfection. It is then no accident when its exercise is a success. To say that there must not be outside interference is not to make success accidental. It is only to make success contingent, to mark it as the achievement of a finite being who acts in an environment not under her complete control among others like her. Whether another kind of being is possible is neither here nor there when it comes to the conditions of agency. The account of perfect rationality needed is an account of perfect human rational agency. After all, the topic is our capacity to act and the basic principles that govern our exercises of it.

Neither God nor any other kind of being has much to do with that.²⁷

27. Lavin elsewhere claims that although “many ... central features of human agency are incompatible with divine perfection”, he does not “want to deny the intelligibility of the idea of divine agency. Nevertheless, whatever intelligibility it has, it has through other channels than giving application to the forms of predication at the ineliminable core of our own intentional agency” (Lavin [2013] 296n.3). Likewise, although central features of human agency are incompatible with other kinds, I need not deny the intelligibility of those other kinds. If they are intelligible, though, we understand them through channels other than self-knowledge of our agency. They would characterize beings utterly unlike us. However we might know about them, such knowledge has no role in an account of our capacity to act.

Let me sum up. The error condition differs from the success condition. Whereas the practical cognition condition transposes the success condition into the key of self-conscious agency, a parallel transposition of the error condition into the willful violation condition is incorrect. Complying with principles marks a perfection of exercises of our capacity to act. Deviating is a characteristic error to which I am susceptible in those exercises that constitutes a failure of them if it is severe enough. The error condition thus uses the notion of what can happen because I have a self-conscious capacity to act, not the notion of what I can do because I have it or of what is possible because I have it. We err because we have incompletely developed capacities to act and because we are not in control of the world and everyone else in it. We most likely will for our entire lives. This statistical claim does not show that perfect human agency is impossible, though, because the idea of an incompletely developed capacity depends on the idea of a completely developed capacity. If I fully develop my capacity to act, I non-accidentally act from my correct understanding of its principle and how it applies to my situation. Absent outside interference, I succeed. Perfect rationality is possible for each of us because we are human beings even if the vagaries of biology, development, our historical, social, and political situation, and just about everything else make its realization in any of us unlikely.

7. OTHERWISE RESPONSIBLE

The Alternate Possibility Condition I am responsible for acting in some way only if I *can* do otherwise.

Because of the way that the metaphysics of capacities orients my understanding of these conditions, this condition is about responsibility for complying with and deviating from the principle of the capacity that

I am exercising. I shall argue that it in fact abstractly characterizes two specific conditions, one about deviance and the other about compliance, that use different notions of possibility:

Condition on Responsibility for Deviance I am responsible for deviating from a principle that governs my action only if I *can* comply with it.

Condition on Responsibility for Compliance I am responsible for complying with a principle that governs my action only if I *can* deviate from it.

I will first argue that I am responsible for deviating from a principle only if I can comply with it, which uses the notion of *what I can do* given the state of development of my capacity to act. I will next argue that I am responsible for complying with a principle only if I can deviate from it, which uses the notion of *what can happen* given the state of development of my capacity to act. The notions of possibility in these conditions differ because they use different subsets of what is possible given my self-conscious capacity to act. The generic alternate possibility condition abstracts away from what explains the difference between the notions of possibility that they use. It is indeterminate and thereby an inapt focus of an inquiry into our agency.

Start with a criticism of my account of the practical cognition condition that says that it cannot capture the sense of internality relevant to basic normative standard for our action. Take the psychopath. Perhaps she is not how she ought to be, but is it not too harsh to say that she does what she ought not to do? She is properly subject to pity, care, and concern. We need to protect us from her and her from herself. Is she also properly subject to blame, resentment, and other attitudes that often go with violating a normative standard on action? It seems not. The futility of explaining to her the authority of this principle seems to cast doubt on that authority. For someone to be subject to a standard that they

cannot appreciate seems to be in some sense unfair or inappropriate. What, then, of the claim that the principle of our capacity to act establishes an internal objective normative standard for her action because she is one of us?

The idea behind this criticism is correct, but it is compatible with my view. It seems otherwise only if you collapse the practical cognition condition and the condition on responsibility for deviance. What is the same thing, it seems otherwise only if you do not separate issues about deviating from a standard in an exercise of a capacity from issues about responsibility for deviating. The practical cognition condition is about whether an agent falls under a principle and thus succeeds by complying with it and fails by deviating from it. Falling under a principle is necessary but insufficient for being responsible for deviating from it. To see why, imagine that a sinkhole opens under me as I walk to the store. I plunge to the depths. I fail in this exercise of my capacity to act. After all, although I was walking to the store, I never walked to the store. I did not finish what I was doing. I am not responsible for this failure, though, because its source is the sinkhole. A sinkhole opening under me is out of my control, ability to predict, and ability to navigate. Roughly, then, I am responsible for deviating from a principle when I am in some sense the source of the failure.

The practical cognition condition concerns success and failure, not responsibility, and the psychopath fails to act as she ought even if she is not responsible for it. After all, when a psychopath acts immorally or imprudently, she is not like the lion or the mosquito who is not in error. She like them does not appreciate moral and prudential requirements on her action. Her ignorance is a failure, though, a result of a defect in the development of her capacity to act. Theirs is not. This, again, is why psychopathy is a kind of pathology, not a different way of being.

Is the psychopath responsible for her failure? Her pathology explains why she does not act on prudential and moral grounds for action. She is thus in some sense the source of her failure. I shall argue that this sense is irrelevant, though, which explains why she is properly the object of pity, care, concern, protection, and quarantine but not other interpersonal responses. Deviating comes apart from responsibility for it because the practical cognition condition and the condition on responsibility for deviance use different notions of possibility. Like the practical cognition condition, the condition on responsibility for deviance uses the notion of what I can do because I have a self-conscious capacity to act, not what is possible or what can happen because I have it. Just try it with either of them. Each implies that I can be responsible for failing to get to the store when a sinkhole opens under me. Neither will do.

The notion of what I can do because I have a self-conscious capacity to act, though, is too broad. It implies that I can be responsible for anything within human abilities, regardless of my actual condition. The problem here is the dual of the earlier problem with the practical cognition condition. That condition must use the notion of what I can do because I have a self-conscious capacity to act independent of my development of it in order to avoid conditioning principles for our action by our imperfections in a way that is incompatible with their objectivity. This condition must instead use the notion of what I can do because I have developed my capacity to act to a particular degree in order to avoid licensing responsibility for failures to act in ways far beyond my current abilities. Otherwise, babies are responsible for an awful lot. I am then responsible for deviating from a principle only if I am able to comply with the principle given the state of development of my capacity to act. With respect to prohibitions, then, this condition says that I am responsible for deviating from a principle by acting in some way only if intentionally refraining from acting in that way is within my current ability. With

respect to requirements, it says that I am responsible for deviating from that principle by not acting in some way only if intentionally acting in that way is within my current ability. In essence, I am responsible for deviating from a principle only if I can comply with it given the state of development of my capacity to act.

This interpretation might seem to let people off the hook too easily. After all, although much goes into developing our capacities, self-cultivation is essential to developing a self-conscious capacity to act. At some point in my life, thinking about the proper grounds of action and habituating myself to act for them when appropriate are as much a part of developing this capacity as the tutelage of others. It might then be my fault that I am not able to comply with the principle of my capacity at some particular time.

However, the description of the problem contains its solution, in two parts. For one thing, developing our capacity to act is in part something that we self-consciously do, at least once we have developed sufficiently. It is an exercise of that very capacity. I can thereby be responsible for failing to adequately develop it in the same way that I can be responsible for any deviant exercise of it. This failure can be the source of other deviations for which I can then be derivatively responsible. That is how to think about someone who cannot now act well because she did not back then develop the skills open to her. For another thing, many actions are temporally extended affairs. Take a case where I take a class with a single exam at the end of a semester. Focus just on the last moment when I sit down to take the test and you might think that I cannot pass it. I lack the knowledge. A course of study, though, takes a while. To focus on the last bit is to miss most of the process. If I lack the knowledge because of lack of effort throughout the semester, I can be responsible for my sorry state of development and the inevitable

failure that results. I have put myself in a position to fail because I did not do enough in order to be in a position to succeed. Like the previous case, I deviate now because of what I did not do over time. Unlike the previous case, though, the deviance here is in one extended exercise of my capacity to act. It is a failure to take enough means early in the action in order to pull off the end when the time comes.

The psychopath is different. She not only cannot recognize moral and prudential requirements now, she also cannot come to recognize them because of her pathological condition. Nor was she ever able to develop her capacity and come to recognize them. This is a kind of characteristic imperfection whose possibility comes with possessing a capacity that develops over time. If we cannot help psychopaths and they cannot help themselves in these areas, we and they lack responsibility all the way down for their sorry state and actions.

In effect, then, I am directly responsible for deviating from a principle only if complying with it is within my ability at the time. I am derivatively responsible only if complying would have been within my ability had I availed myself of the opportunities to develop that capacity. That is to say that I am directly responsible for deviating from a principle only if I can exercise that capacity in compliance with the principle given my state of development of that capacity. I am derivatively responsible only if I would have been able to exercise the capacity in compliance with the principle given the state of development of that capacity that I would have possessed had I availed myself of opportunities to develop it. Such is the proper interpretation of the condition on responsibility for deviance that uses the notion of what I can do given the state of development of my capacity to act. In effect, then, the notion of possibility in the condition on responsibility for deviance relativizes the notion of possibility in the practical cognition condition. Whereas the practical cognition condition uses the notion of what I can do because I have a

self-conscious capacity to act, the condition on responsibility for deviance uses the notion of what I can do given the state of development of that capacity.

Whereas the condition on responsibility for deviance uses the notion of what I can do given the state of development of my capacity to act, the condition on responsibility for compliance uses the notion of what can happen given that state of development. This difference between the conditions is straightforward. When I deviate, I fail in my exercise of that capacity and thus manifest an element of the 'what can happen' subset of what is possible given that I possess that capacity. The question about responsibility then turns on whether actualizing an element of the other subset in those circumstances is possible. After all, whether other ways of deviating are available is neither here nor there when the question is whether I am responsible for deviating from a principle. If all that is possible is that I deviate and the only difference that I make is in which way, I cannot be responsible for deviating. The crucial question is whether I can comply. When I comply, I instead succeed in my exercise of that capacity and thus actualize an element of the 'what I can do' subset of what is possible given that I possess that capacity. The question about responsibility then turns on whether manifesting an element of the other subset in those circumstances is possible. After all, whether other ways of complying are available is neither here nor there when the question is whether I am responsible for complying with a principle. If all that is possible is that I comply and the only difference that I make is in which way, I cannot be responsible for complying. The crucial question is whether I can deviate.

Hence, the two conditions use different notions of possibility, with the one about deviance using the notion of compliance or what I can do and the one about compliance using the notion of deviance or what can happen. Yet both of these conditions relativize the subset of what is possible given my capacity

to act to my state of development of that capacity. This capacity develops over time, and I come to be able to do things that I previously could not do. Since the question about responsibility for deviance is about whether I can comply with the principle in question, the condition needs to use the notion of what I can do relative to the state of development of my capacity to act. Otherwise, holding me responsible would be unfair. Yet with the development of that capacity and thus the increase in my abilities brings with it possibility of interferences, interruptions, overreaches, and other kinds of errors that constitute deviance if they are severe enough. An agent with a completely undeveloped capacity to act cannot exercise it at all, let alone succeed or fail in exercising it. An infant cannot trip until she can walk. An agent with a developed capacity to act can exercise it and thus can succeed or fail in exercising it and succeed or fail to exercise it. Only a human who develops a bit can walk, and only then is success and failure in walking and success or failure to walk possible.

With the development of a capacity to act come two kinds of defects, internal ones wherein an incompleteness of the development leads to error and external ones in which something outside of the capacity does. For a while, the increase in abilities brings with it an increase in both kinds of defects. Only a being who can walk can misstep, and only a being who can walk can be run down by a truck in mid-stride. A capacity develops along a spectrum, though, and as it tends toward perfection, possible internal defects decrease. A perfectly developed capacity issues in no exercises with internal imperfections, though it still issues in exercises with external imperfections given that human beings are finite animals in a world not under their will. Hence, the state of development is essential to the notion of possibility in the condition on responsibility for compliance as well as the notion of possibility in the condition on responsibility for deviance. Only with it does the condition get the scope of the possibilities

for deviance that an agent avoids in complying correct.

The notions of possibility in both conditions on responsibility are thus relative to the state of development of my capacity to act. They differ because the condition on responsibility for deviance uses the notion of what I can do given my state of development of that capacity whereas the condition on responsibility for compliance uses the notion of what can happen given my state of development of that capacity. Since the generic alternate possibility condition abstracts away from these distinctions, it obscures the notions of possibility needed in order to understand these conditions on responsibility. It is thereby an unfit object of inquiry.

In order to explain that last point, let me highlight one important implication of this account of the notions of possibility in these conditions. How does the practical cognition condition relate to the alternate possibility condition? Since the alternate possibility condition is in fact a generic version of two prior specific conditions, this question breaks up into two questions. One is about the relationship between the practical cognition condition and the condition on responsibility for deviating. The other is about the relationship between the practical cognition condition and the condition on responsibility for complying. The literature on this issue in fact focuses on the first question and asks whether the practical cognition condition—under the name *ought implies can*—entails the condition on responsibility for deviance. That is to say, the question is about the soundness of the following hypothetical syllogism:

If I ought to ϕ , I can ϕ .

If I can ϕ , I can be responsible for failing to ϕ —for deviating.

Therefore, if I ought to ϕ , I can be responsible for deviating.

On my view, the proper interpretations of the notions of possibility in these conditions differ. Whereas both use the notion of what I can do because I have a self-conscious capacity to act, the second

relativizes it to my state of development of that capacity. The other does not and instead represents that capacity as I possess it because I am a human being. This inference thus equivocates over the 'can's in the consequent of the first conditional and the antecedent of the second.

An equivocation, though, might only show that the inference is enthymematic, missing a premise linking the 'I can ϕ 's in the first and second premises. However, the proper interpretations of the conditions in fact renders the inference unsound. In particular, some principles meet the practical cognition condition with respect to some agent and action even though they do not meet the condition on responsibility for deviating. Such is the case with the psychopath. Hence, nothing follows about whether a principle meets the condition on responsibility for deviating with respect to an agent and action from the fact that it meets the practical cognition condition with respect to them. Correlatively, nothing follows about whether it meets the practical cognition condition with respect to them from the fact that it does not meet the condition on responsibility for deviating with respect to them. However, if it does not meet the practical cognition condition with respect to them, then it does not meet the condition on responsibility for deviating with respect to them. After all, if I in principle cannot comply with the standard, I cannot comply with it no matter the state of development of my capacity to act. If it meets the condition on responsibility for deviating with respect to them, then it meets the practical cognition condition with respect to them. If I can comply with a standard given the state of development of my capacity to act, I can in principle comply with it.

What about the relationship between the practical cognition condition and the condition on responsibility for success? Here the first challenge is to formulate a clear connection between the following two conditionals.

If I ought to φ , I can φ .

If I can fail to φ , I can be responsible for φ -ing—~~for complying~~.

After all, the condition on responsibility for deviance links the actuality of success— φ -ing, in this case—with the possibility of error—failing to φ . The notions of possibility are thereby explicitly distinct in a way that does not allow for even the appearance of a simple hypothetical syllogism connecting the two conditions. Once again, auxiliary premises must connect them. The easiest route is with a premise linking the consequent of the first condition with the antecedent of the second.

If I ought to φ , I can φ .

If I can φ , I can fail to φ .

If I can fail to φ , I can be responsible for φ -ing—~~for complying~~.

If I ought to φ , I can be responsible for complying.

Yet the second premise of this argument is ambiguous. The consequent of the first conditional uses the notion of what I can do given my capacity to act as I possess because I am a human being. The antecedent of the third conditional uses the notion of what can happen given the state of development of my capacity to act. The ambiguity of second conditional is clear if you ask whether you start interpreting it with by factoring in the notion of possibility in the antecedent of the consequence of the first conditional or the antecedent of the third. Start the first way and the *I can fail to φ* in the second conditional will not be relative to the state of development of my capacity to act. After all, what comes with possessing a capacity by nature are the kinds of error to which I am by nature subject, not the subset of those errors to which I am subject given my state of development. Start instead the second way and the *I can φ* in the second conditional will be relativized to the state of development of my capacity to act. After all, the subset of errors that are relativized to the state of development of my capacity to act come with the subset of abilities that are relative to that state of development. Hence, no reading of the second

conditional is available in order to render the inference valid.

Moreover, again, even with more auxiliary premises, any such inference between these conditions is invalid given the correct interpretations of the practical cognition condition and the condition on responsibility for complying. In particular, some principles meet the practical cognition condition with respect to some agent and action even though they do not meet the condition on responsibility for complying. Although the principle may govern my capacity, my compliance might be guaranteed by factors outside of my control such that I cannot deviate and thus cannot be responsible for complying. Hence, nothing follows about whether a principle meets the condition on responsibility for complying with respect to an agent and action from the fact that it meets the practical cognition condition with respect to them. Correlatively, nothing follows about whether a principle meets the practical cognition condition with respect to an agent and action from the fact that it does not meet the condition on responsibility for complying with respect to them. However, if a principle does not meet the practical cognition condition with respect to an agent and action, it does not meet the condition on responsibility for complying with respect to them. If a principle cannot govern my action, I cannot be responsible for complying with it because of the way that constitutivism links compliance with principles governing my action. There is no way to comply with such a principle and hence no way to be responsible for complying. Moreover, if a principle meets the condition on condition on responsibility for complying with respect to an agent and action, it meets the practical cognition condition with respect to them. If I can be responsible for complying, the principle must govern my exercises of my capacity to act.

Let me leave off this discussion and summarize this section. The interpretations of these

conditions and the claims about the relationships between them and the practical cognition condition are controversial. The conditions themselves are only necessary conditions on responsibility, not sufficient conditions, for responsibility, let alone for blameworthiness for deviance or praiseworthiness for compliance. Still, the important point for present purposes is that the two conditions that establish the generic alternate possibility condition use notions of possibility relative to the state of development of my capacity to act. They thereby differ from the practical cognition and error conditions. Each condition on responsibility takes that for which I might be responsible—deviance in the one case, compliance the other—and matches it with the possibility of its opposite pair—the possibility of compliance given the state of development of my capacity to act in the one case, the possibility of deviance given that state of development in the other. I can be responsible for deviating if complying is within my ability, and I can be responsible for complying if deviating can happen given my abilities. Such are the proper interpretations of the conditions on responsibility for deviating and complying.

8. CAPACITIES UNQUALIFIED

Whereas the conditions of agency use modal vocabulary, the theses about our agency do not:

The Guise of The Good To act is to act in a way that I take to be good.

The Action-Knowledge Link To act intentionally is to act knowingly

Practical Reason Is The Will To judge practically is to decide.

Modal vocabulary instead appears in summary objections based on putative counterexamples:

The Guise of The Good (Modal Version) Necessarily, a human being acts in some way only if she takes acting in that way to be good.

Objection: Possibly, a human being acts in a way that she does not take to be good.

The Action-Knowledge Link (Modal Version) Necessarily, a human being acts intentionally if and only if she acts knowingly.

Objection: Possibly, a human being acts intentionally without knowing what she is doing, and possibly, a human being acts intentionally without knowing why she is doing it.

Practical Reason Is The Will (Modal Version) Necessarily, a human being judges that she should ϕ if and only if she decides to ϕ .

Objection: Possibly, a human being judges that she should ϕ without deciding to ϕ , and possibly, a human being decides to ψ without judging that she should ψ .

Philosophers tend to take the unqualified nature of the theses to indicate that they are claims meant to hold with necessity. They treat them as universal generalizations about exercises of our capacity to act. Defenders of these theses tend to respond to the objections by accepting this characterization of them while disputing the interpretation of the examples. For example, defenders of the guise of the good thesis often claim that humans who knowingly act wickedly take the wickedness to be a good-making feature of the action. Maybe it is sexy, and being sexy is good. Another counterexample is next in line. An interpretation that finds something good in the action from the agent's point of view follows in lockstep. The dialectic seems interminable. The defense strains credulity and dilutes the thesis to the point that it loses interest or importance. The thesis in the end becomes hard to deny, partly because nothing can count against it. It is thereby unworthy of philosophical attention.

Critics of the action-knowledge link point to actions where I do not know what I am doing because of an epistemic limit and actions where I do not know why I am acting because of a lack of self-knowledge. Maybe I do not know whether I am signing the ten carbon copies even though the correct description of my action is signing ten carbon copies in order to authorize the contract. There are a lot of them. I doubt I can press hard enough to ink them all, and the mobster is only giving me one go, no checking or retracing. Maybe I do not know why I am being mean to you because my envy is repressed

beyond even the talents of my analyst. Standard responses grant the characterization of the thesis and dispute the interpretation of the examples. They tend to retreat to easier or more trivial self-knowledge, perhaps that I am pressing hard in order to sign ten copies and that I am being mean to you in order to express some desire or emotion or other, who knows which. At each point, a case arises that challenges even this knowledge, and the retreat proceeds to a weaker claim. At some point, the knowledge is so restricted that it is unclear what it has to do with my action—that particular material process in the world. Even then, claims to self-knowledge are fallible. Onward the retreat beats until it is unclear what it has to do with action and why anyone should care.

Critics of the thesis that practical reason is our kind of will point to cases of *akrasia* wherein I judge that I should act in some way but decide to act otherwise. Similarly, cases of rashness show that I can decide to act in some way without thinking a lick about how I should act. Defenders of the thesis tend to deny the interpretations of the cases. Socrates, according to legend and Plato, doubts the possibility of *akrasia*, and Aristotle, on a standard reading, claims that *akrasia* is possible only if the judgment is inaccessible to the agent at the moment of decision. Similarly, someone might claim that every decision is identical to a judgment even if I do not take myself to have a view on what I should do. After all, perhaps this decision embodies my implicit judgment that I should decide to act in this instance without thinking too hard about the matter. Spontaneity and all that. Again, more counterexamples arise, and defenses that weaken the thesis follow. Even if they work, they take the life out of it.

I am not doing justice to any of these debates. My interest here, though, is in the road not taken. Defenders of the theses assume that they must defend the theses *as the critics understand them*. This assumption leads down the beaten path. I do not think that these responses work. More importantly, I

do not think that it matters. The theses are not claims meant to hold with necessity. They are not about every exercise of our capacity to act. They instead describe its nature and are thereby normative for its exercises. This understanding neutralizes the putative counterexamples. In fact, it shows that they are evidence for the theses that they are meant to refute. So I shall now argue.

Start with the guise of the good thesis. If it is about every exercise of our capacity to act, it is incompatible with the existence of evil, at least on a certain interpretation. On this interpretation, when I act self-destructively, I do not act in ways that I know to be bad, despite what I might say in the midst of my depression. I instead act in ways that I take to be good, with the explanation likely picking out as good some kind of disobedience or denial of things commonly taken to be good. While this kind of explanation might be plausible in some cases, to offer it in every case seems like the kind of thing that only comes from commitment to a theory. In particular, it seems like it comes from a commitment to how our sense of the good must relate to our action if goodness stands to our action differently than badness does. However, the metaphysics of capacities explains how this distinction is possible without the implausible modal interpretation of the thesis.

The basic thought behind the thesis is that goodness in some sense characterizes our action. In particular, goodness in some way characterizes the thought that we realize in action. To characterize our thought and action, though, it does not need to be true of every exercise. It instead needs to characterize the nature of the capacity and thus be a perfection of exercises of that capacity. To not act in a way that I take to be good is a defect or imperfection in an exercise that constitutes a failure of that exercise if it is severe enough. Defenders of the guise of the good thesis thus need not deny the possibility of action under the guise of the indifferent or the guise of the bad. In fact, such a denial is incompatible with the

proper understanding of the thesis. After all, anything that describes the nature of our capacity to act marks a perfection of every exercise of that capacity. With perfection comes imperfection. Characteristic imperfections can happen in an exercise of that capacity. They derive from the nature of what I can do because I have this capacity, marking possible privations in exercises.

The putative counterexamples to the thesis are then in fact evidence for it. Here is what I mean. According to this interpretation, to act under the guise of the bad or the guise of the indifferent is a per se defect in that exercise of my capacity to act. To the extent that acting in these ways strikes us to that extent imperfect or defective, these cases are evidence for the guise of the good thesis. Since they are imperfections, their complement must be a perfection. For their complements to be a perfection is for acting under the guise of the good to be part of the nature of our capacity to act and thereby normative for its exercises. The guise of the good thesis thus explains the possibility of not acting under it and the imperfection therein. It explains why action that is under the guise of the good and action that is not are possible while establishing an asymmetry between them that shows that one is a perfection, the other an imperfection, in exercises of our capacity to act.

Moreover, this interpretation of the thesis is more substantial than the modal interpretation. Good exercises of our capacity to act meet its principle. Because I act self-consciously, a perfect exercise of my capacity to act includes knowingly acting from my understanding of what follows from the principle that governs this capacity. To act under the guise of the good, then, is not to act simply in a way that I take to be 'good' in some thin sense of this term. It is to act in a way that meets the basic normative standard for exercises of this capacity from an understanding of that standard and its bearing on my action. To act under the guise of the good is to act under the guise of the actual substantive good for

exercises of this capacity. If its principle includes moral, prudential, and rational content or something that unifies them, to act under the guise of the good is in part to act with a sense of the moral, prudential, and rational standing of acting this way. To the extent that acting without regard for the moral, prudential, and rational standing of my action is to that extent imperfect, it is evidence for the guise of the good thesis on its proper interpretation.

The other theses likewise describe the nature of our capacity to act and thereby are normative for its exercises on the proper interpretation of them. Since the structure of the explanation is identical, I shall be quick. Turn now to the action-knowledge link. If the thesis describes the nature of our capacity to act, exercises that deviate from it are not counterexamples to it. They are evidence for it. To act knowingly marks out a perfection of its exercises, and not knowing what I am doing and why I am doing it are characteristic imperfections in those exercises. Everyone thinks that these kinds of ignorance or mistakes are imperfections in our exercises of that capacity. That is why they are ignorance and mistakes. Hence, this interpretation of the condition explains why these deviant cases are possible and also why they are deviant. Unlike the defenders of the modal interpretation, it does not and indeed cannot deny their possibility. Unlike the critics of the modal interpretation, it does not and indeed cannot treat cases of ignorance and cases of knowledge as on a par. It explains why both are possible while establishing an asymmetry between them in a way that shows that one is a perfection, the other an imperfection, in exercises of our capacity to act.

Likewise for the thesis that practical reason is our kind of will.²⁸ As a modal claim, the idea that

²⁸ This thesis is a more general version of the metaethical thesis of motivational or judgment internalism about moral judgment. My sense is that the view about moral judgment derives its plausibility from this general view about practical judgment and the assumption that a moral judgment is always an (all-out) practical judgment.

judgment and decision mark two aspects of one act is implausible. If the thesis instead describes the nature of our capacity to act, though, exercises that deviate from it are not counterexamples to it. An exercise in which I am akratic or otherwise defy this thesis is to that extent imperfect. Since akrasia is a breakdown or failure in our agency if it is anything, its possibility is not in tension with the thesis. Its possibility in fact follows from the thesis in a way that explains why it is an imperfection or defect. In this way, instead of being counterexamples, exercises of my capacity to act that violate the thesis are evidence for it. Because akrasia is a per se defect in an exercise of my capacity to act, acting in a way that realizes the thesis is a per se perfection. Since the principle that describes the nature of a capacity establishes a normative standard for its exercises, a per se perfection of an exercise is the realization of a feature that describes the nature of our capacity. Unlike the defenders of the modal interpretation, this interpretation does not and indeed cannot deny the possibility that judging and deciding can come apart. Unlike the critics of the modal interpretation, it does not and indeed cannot treat cases in which they do not come apart as on a par with cases where do. It explains why both are possible while establishing an asymmetry between them in a way that shows that one is a perfection, the other an imperfection, in exercises of our capacity to act.²⁹

The metaphysics of capacities thereby interprets these theses as interesting and true. It offers the explanatory structure needed to explain why deviant exercises of this capacity are possible, both in the sense that they can happen and in the sense that they are deviations or per se imperfections. In doing so, it shows that putative counterexamples to the theses are in fact evidence for them. It thus vindicates the

²⁹ This argument does not establish that practical reason is the will. For all it shows, practical reason and the will might be distinct capacities with an essential relationship to each other. I argue in other work against separating practical reason and the will and in favor of the view that practical reason is the will.

theses without resorting to desperate and implausible reinterpretations that characterize defenses of the modal interpretations and without making them uninteresting and unrecognizable. Of course, each thesis deserves more discussion than I give it here. The important point, though, is the structure of explanation that the metaphysics of capacities makes available and the way that it alters the debate about the theses. To characterize a capacity is not to be true about every exercise of it. It is to be true of every perfect exercise of it.

9. CONSTITUTIVISM AND 'CAN'

Although a characteristic 'can' often pops up in practical philosophy, it expresses different notions in different contexts.

The Success Condition A principle governs my action only if I *can* comply with it—that is, I *can do* what it requires and permits and I *can refrain from doing* what it prohibits given my capacity to act as I by nature possess it.

The Error Condition A principle governs my action only if I *can* deviate from it—that is, only if what it prohibits *can happen* given my capacity to act as I by nature possess it.

The Practical Cognition Condition A principle governs my action only if I *can* self-consciously comply with it—that is, only if I *can do* what it requires and permits and only if I *can refrain from doing* what it prohibits from my understanding of what it requires, permits, and prohibits given my capacity to act as I by nature possess it.

The Alternate Possibility Condition I am responsible for acting in some way only if I *can do* otherwise, possibly deviating from a principle with which I in fact comply and possibly complying with a principle from which I in fact deviate. When I comply, I am responsible for it only if what the principle prohibits *can happen* given the state of development of my capacity to act. When I deviate, I am responsible for it only if I *can do* what the principle requires given the state of development of my capacity to act.

A metaphysics of capacities explains what unifies these notions of practical possibility and what distinguishes them from each other. We thereby understand how the possibility of error comes with the possibility of success even though success concerns what I can do and error concerns what can happen

given what I can do. We thereby understand how compliance and deviance can come apart from responsibility for complying and deviating. We thereby understand how aspects of our agency that do not describe every exercise of it mark perfections of those exercises because they characterize the capacity. We must live up to them because of what we are even though we do not always live up to them because of what we are. That is what it is for the principle of our capacity to act to establish an internal objective normative standard for our development and exercises of that capacity. What I am determines how I am to be and what I am to do.

INTELLECTUAL ISOLATION

All this intellectualist legend must be rejected, not merely because it tells psychological myths but because the myths are not of the right type to account for the facts which they are invented to explain. (Ryle [1946] 228)

1. THOUGHT AND ACTION

'HOW AM I TO ACT?' As far as we know, only we can ask this question and act from our answers to it. My cat can act in ways that are good or bad for others and for himself. I can act in some way, though, *because* it is good for others and for myself. In this way, only human beings have a self-conscious capacity to act.³⁰

Take these statements of this idea by theorists whose account of it I shall criticize:

[A]ctions of rational agents are guided by and responsive to their deliberative reflection about what they have reason to do. (Wallace [1999] 44)

[A]gent's rational capacities enable them (fallibly) to identify some values in some options and to respond to them, i.e. to recognize that those aspects of the option that make it valuable are reasons for taking it, and they enable them also to do so, to take that option for that reason. (Raz [2011] 4)

Our practical reasoning should not end with ... normative beliefs. To be fully practically rational, we must also respond to practical reasons or apparent reasons with our ... acts. (Parfit [2011] II.424)

A rational agent ... is capable of thinking about reasons for certain actions ..., and ... reaching conclusions about which of these are good reasons ... [that] make a difference to the actions ... that it proceeds to [perform]. ... [I]f a rational agent believes that *p* is a conclusive reason to do *a*, she generally will do *a*, and do it *for this reason*. (Scanlon [2014] 54)

These theorists identify two aspects of our practical thought. First, such thought is subject to a standard of correctness. For instance, I am incorrect if I think that I am to belt 'I Put A Spell on You' in the middle of a colloquium. I am correct if I think that I am to save the Nina for departmental karaoke. Second, I can

30. Nothing turns on the terminology of *capacities*. Powers, faculties, dispositions, abilities, capabilities, kinds of activities, and any other term or phrase that similarly distinguishes between a kind and instances or exercises of it are fine.

act from such thought or self-consciously do what I represent as what I am to do. If I think that I am to ask my lover for menu suggestions for our soiree, I can thereby put the question to him.

Putting these aspects together, we can act from thought that is subject to a standard of correctness, which distinguishes us from the other animals. This basic idea allows for many interpretations. In this paper, I will argue that many philosophers explain the standard of correctness of practical thought in a way that is incompatible with acting from that thought. I call this explanation of the standard of correctness *intellectualism*.³¹ As I like to put it, intellectualism is the view that practical reason is a species of theoretical reason, distinguished from others by its objects: reasons to act.³² The basic idea behind intellectualism is that to exercise reason is to represent an object that exists independently of my thought about it, and I represent it well only if I non-accidentally represent it accurately. Just as a tree and the number 7 exist independently of my thought about them, so a reason for me to act exists independently of my thought about it. To exercise reason well with respect to any of them is to non-accidentally represent them accurately. The object and the content of the representation of it make a judgment a practical judgment—an exercise of that species of theoretical reason, not another.

Intellectualism is compatible with many accounts of the *constitution* of reasons to act. They might be fundamental or might depend on other normative or non-normative stuff. They might be ‘objective’ or ‘subjective’. They might be internal or external in the senses familiar from the

31. I take the name from Lavin [2012].

32. Intellectualists might claim that there is just reason, with theoretical reason and practical reason as species distinguished by their contents. However, as I will explain, intellectualists think of reason as such as philosophers who distinguish theoretical reason and practical reason on formal grounds think of theoretical reason. My formulation is preferable since it does not define the alternative out of existence.

internal/external reasons debate. They might be 'ontic' or 'non-ontic' facts, 'substantial' or 'insubstantial' properties, 'world-involving' or 'non-world-involving' truths. Some of them might exist or error theorists might be right. And so on for many metaethical debates. Intellectualism is compatible with these views about the constitution of reasons to act because it is about how I relate to reasons to act in thought, not about those objects of thought. It is about the form of the representation that constitutes an exercise of practical reason and its relationship to its object.

An argument against intellectualism thus has much broader application than many familiar criticisms in practical philosophy. For example, familiar metaphysical and epistemological objections to a certain kind of objectivism about reasons to act only target intellectualist views with specific accounts of the constitution of reasons to act. Likewise, certain arguments against a certain kind of subjectivism about reasons to act only target intellectualists views with specific accounts of the constitution of reasons to act. Even if these arguments work, they leave the basic picture of practical reason intact and only challenge one way of working within that framework. My criticism instead addresses what unifies intellectualists of various stripes and does not depend on the details that differentiate them.

I will show that this intellectualism entails a 'compositional' account of practical thought, by which I mean it decomposes practical thought into two the exercises of two capacities, practical reason and the will. I shall argue that if practical reason and the will are distinct capacities, I cannot act from my practical judgment. I will also argue that if I cannot act from practical judgment, intellectualism is incompatible with a certain condition on the possibility of reasons to act. Without reasons, though, a species of theoretical reason about them cannot exist. Intellectualism is thus self-undermining.

Before I start explaining intellectualism and why philosophers go in for it, let me state my true

aim in this paper. I wish to clear ground for a view that says that practical reason differs from theoretical reason in *form* rather than merely in the content of its representations. It says that practical cognition or thought is a different *kind* of cognition or thought than theoretical cognition or thought. I find this kind of *practical cognitivism*, as we might call it, in Aristotle, Immanuel Kant, Elizabeth Anscombe, and Christine Korsgaard, among others, though interpretations vary.³³ Regardless of who holds the view, though, these ideas are not much on the scene these days and are difficult to state, let alone defend, within the theoretical confines of contemporary practical philosophy. Just think about how ‘theoretical cognition’ or ‘theoretical thought’ sound like pleonasms. Think about what that implies about how we hear ‘practical cognition’ and ‘practical thought’. Intellectualism is the default position, not because we consider and reject the alternative but because we think about things in a way that keeps us from considering it. My goal is to focus on it and expose it in order to let us start thinking about and developing an alternative.

2. PRACTICAL REASON AND THE WILL

Joseph Raz claims that “practical reasoning is ... but reasoning about a particular subject matter ... [;] there is no distinctive form of reasoning called practical reasoning” (Raz [2011] 139). He here contrasts two views, the intellectualist view and a view that distinguishes practical reason and theoretical reason ‘formally’. We can understand the explanatory structure and burdens of intellectualism only if we understand the alternative view. To that end, I start with some issues that intellectualists take to show that practical reason and the will are distinct capacities. This separation is the crux of their account of

33. I take the name from Engstrom [2009]. In light of it, we can call intellectualists ‘theoretical cognitivists’. These labels have the virtue of not assuming that all cognition is theoretical cognition and all objects of reason are objects of theoretical reason.

practical thought. To oppose intellectualism is to deny it.

Like many intellectualists, R. Jay Wallace claims that Buridan cases and cases of akrasia show that practical reason and the will are distinct capacities even though often there “is no phenomenological gap between ... normative judgment ... and ... intention” (Wallace [2001] 97). Say I come to two grassy roads wanting wear that diverge in a yellow wood. They lead to the same place, are equally long, are equally covered in leaves that no step has trodden black, and all the rest of it. Everything against or in favor of taking one route is against or in favor of taking the other. Although I can represent a reason to take either road, then, I cannot represent a reason to take one over the other. Just try it. I therefore cannot form a judgment that I should take one over the other. If all there is to human agency is the ability to make judgments about reasons to act, I will never act. I will just sit here at the fork in the road and probably die like that donkey, at least if nothing relevant changes. Of course, I can take one of the roads, and hence there is more to our agency than the ability to make judgments about reasons to act. According to intellectualism, this something more is our will, which is separate from practical reason. Through it, I can choose between actions that I cannot distinguish on normative grounds through practical reason. I can self-consciously act even in Buridan cases because I have a will in addition to possessing practical reason.

Separating practical reason and the will also seems to explain the possibility of akrasia. Say I judge that should work on my paper but instead hit the dance floor. How is this conflict possible? Intellectualists think that the answer falls out of a fuller description of the case. Properly speaking, I judge that I should do something but decide to do something else. Because I act from decision, not judgment, I act as I decide, and akrasia is possible.

After reflecting on these kinds of cases, Wallace concludes that we “have the capacity for a sophisticated kind of rational agency, insofar as [we] can reach independent normative conclusions about what [we] have reason to do, and then choose in accordance with such normative conclusions. ... Once we have this power, however, it can be put to use in ways that are at odds with our own practical judgments” (ibid. 92). On the same basis, Derek Parfit says that when “we conclude that we ought to do something, we are not deciding to do this thing, but coming to have a normative belief. ... We always have two questions: Q1: What ought I to do?; Q2: What shall I do?” (Parfit [2011] II.386). Intellectualists thus claim that practical thought “involves making up my mind twice. Making up my mind about what is the best thing to do is a judgment Making up my mind about what to do is forming an intention” (Watson [2003] 128). Hence, these cases reveal a distinction between judgment and decision that is always present even though our phenomenology does not always present it.³⁴ To judge is not to decide, to decide not to judge. Practical reason and the will are distinct capacities.

To separate practical reason and the will is to decompose practical thought. To deny intellectualism is to deny this separation and instead say that “the will is nothing other than practical reason” (Kant [1785] 4:412). This statement is too stark, though, because animals who lack practical reason have wills and act. Better said, the alternative is that practical reason is our kind of will. It is the

34. Many endorse this line of reasoning, sometimes surprisingly. For example, Allan Gibbard claims that “*thinking what I ought to do is thinking what to do*” (Gibbard [2003] ix-x). He also says that “conclusions on *what to do* have an automatic bearing on what to do. ... [D]eciding ... is not ... coming to a belief in some special kind of fact” (ibid. 5). According to him, then, practical reason is the will. However, in that work, he does not explain how I conclude what to do. He tries to fill this gap in *Reconciling Our Aims* when he says that I decide on the basis of ‘intuitions of the good’. From this idea, he concludes that practical thought has “two stages: In the first stage, I form my valences and preferences. In the second stage, if there’s more than one thing I equally and most prefer from among my alternatives, I pick one—not out of preference, but out of the necessity to choose if I’m not to be like Buridan’s ass. My strictly normative thinking is a matter of the first stage. ... Thinking what I ought to do, then, is not all of thinking what to do” (Gibbard [2008] 19). To form valences and preferences, though, is to ‘come to a belief’ about which action has a particular property. It never involves a decision, even if I most prefer one thing. On this view, then, Buridan cases show that practical reason and the will are distinct capacities.

self-conscious will. Practical reason and the will are thus the same capacity *in us* even though other animals have other kinds of wills.

One way to get into this alternative to intellectualism is to consider Anscombe's distinction between theoretical and practical knowledge, where they are the perfect exercises of theoretical and practical reason that reveal the natures of the capacities. Theoretical reason first, practical reason after.

When I exercise theoretical reason, I represent objects that do not depend on that representation. I represent them as independent of that representation. A tree, say, is what it is regardless of my thought about it. It is an 'independent existence', and this fact about the way that I relate to the object also characterizes my representation of it. This aspect of my relationship to the object is, as it were, a formal part of that representation, not part of its content. It is the way that I represent the object, and I can articulate it by reflecting on the exercise. Theoretical reason is a capacity to represent independent existences as such. I thereby must represent the world as it is in order to exercise theoretical reason well. To exercise it well, then, is to non-accidentally represent an independent existence accurately. Because the object is an independent existence, I represent it *accurately* when I represent it as it is independent of my representation of it. I *non-accidentally* represent it accurately when I stand in a proper relationship to it like, say, a perceptual or testimonial relationship. The basic question of theoretical reason is then 'Whether *P*?' or, generally, 'What is the case?' To answer it is to have "speculative knowledge" that "is derived from the objects known" (Anscombe [1957] §48).³⁵

35. The role of first-person in these accounts of theoretical and practical reason is important. Although theoretical knowledge is knowledge of something as independent of *this* representation of it, it is not thereby knowledge of something as independent of *any* representation of it. If George Berkeley is right, every object depends on God's representation of it. Still, he need not reject my account of theoretical and practical reason. He need only claim that God's (primary) relationship to an object is not theoretical, which is what the devout should about a being who is in no way passive or receptive. More terrestrially, although Kant thinks that all possible objects of theoretical knowledge for human beings in some sense depend on the sensible and intellectual aspects of our capacity to know them, he does not think that a tree depends on *my* theoretical

Practical reason differs theoretical reason in those respects. When I exercise it, I represent objects that depend on my representation of them. I represent them as depending on that representation. For example, I represent cooking with my sweetie in order to spend time with him, and I spend time with him by cooking with him because of that representation. The object—the action—by nature depends on my representation of it in that way. That is why I am not acting if I sleepwalk through similar bodily movements. The object is a ‘dependent existence’, and this aspect of my relationship to the object characterizes my representation of the object. This aspect of my relation to the object is, as it were, a formal part of that representation. It is the way that I represent the object, and I can articulate it by reflecting on the exercise. Unlike the object of a theoretical representation, though, this fact is also part of the content of a practical representation. That content consists of a dependent existence, which is an object that depends on that very representation of it. I thereby must realize the object of my representation in order to exercise practical reason well. To exercise it well, then, is to non-accidentally represent a dependent existence accurately. I *non-accidentally* represent this object when I act in some way because doing so realizes this object. I non-accidentally represent this object *accurately* when the true description of my action as an intentional action is a determinate version of the content of my representation. The basic question of practical reason is then ‘Am I to ϕ ?’ or, generally, ‘How am I to act?’ To answer it is to have “practical knowledge” that is “the cause of what it understands” or ‘of objects that derive from this knowledge’ (ibid.).³⁶

representation of it. After all, you can know the same tree as independent of my representation of it or, for that matter, your representation of it. Finally, I can have theoretical knowledge of your action or, for that matter, my action. Although the action is independent of that theoretical representation of it, it is not thereby independent of the practical representation of the agent in question.

36. ‘What is the case?’ and ‘How am I to act?’ are the *basic* questions of theoretical and practical reason, not the *only* questions that I can address in exercises of these capacities. I can ask not only ‘What is the case?’ but ‘What would have been the case?’, can think not only about real objects but also imaginary objects. I can ask not only ‘How am I to act?’ but also ‘How would I

As we might put the basic idea, a theoretical representation *describes* its object while a practical representation *produces* its object.³⁷ A theoretical representation answers to its object in the sense that the tree, say, is as it is and my job is to represent it as such. The object of a practical representation answers to that representation in the sense that my decision to walk to the store, say, determines whether my subsequent movements are proper realizations of that representation. As we might put it, my job in a theoretical representation is to represent an object that is actual, my job in a practical representation to make actual an object that I represent. If this difference between representations maps onto the difference between theoretical reason and practical reason, practical reason is our kind of will. It is our capacity to determine ourselves to act.

Since intellectualists think that practical reason is not the will, they need a different account of it.

have to act?', can think not only about actually acting but also about hypothetically acting. Still, an account of the nature of the capacity focuses on them in order to understand the function of the capacity in our lives, and such an account must not stick to what the factual and counterfactual, actual and hypothetical, share. Otherwise, practical reason has no intrinsic relationship to acting, nor does theoretical reason have an essential connection to knowing and understanding our world.

37. This way of talking resembles the view that differentiates theoretical and practical reason in terms of propositional attitudes with different directions of fit. In fact, many philosophers think that Anscombe's discussion of the different ways that a list can relate to the items in a shopping cart is the source of the direction-of-fit view (Anscombe [1957] §32). Although I think that this reading reveals more about contemporary philosophical preconceptions than it does about her thought, I cannot defend her honor here. I can, though, briefly explain why I think that the direction of fit view is an inadequate account of the difference between theoretical and practical representations, at least as I have here explained them drawing on Anscombe.

According to the direction-of-fit view, a theoretical representation consists of a mental state with mind-to-world direction of fit like a belief, and a practical representation consists of a mental state with world-to-mind direction of fit like a desire. The view says that these states share a single kind of relationship to an object, though the direction of the relationship flips. Since an object is independent of a belief about it, then, it must be independent of a desire about it in the relevant sense. A desire might cause its object whereas an object causes the belief about it, or mismatch between a desire and its object might indict the object whereas mismatch between a belief and its object indicts the belief. Regardless of whether a direction-of-fit theorist prefers a causal or normative version, or any other version, still, metaphysically, the object of a desire is not *essentially* an object of that desire. The object of a practical representation then cannot be a dependent existence because the object of the belief is not. This is, in effect, to deny the practicality of a practical representation as characterized by the essential interdependence of a practical representation and its object. Yet if you stick with that practicality, you cannot offer an account of a single relationship that theoretical and practical representations bear to the same object in different directions. A representation of an independent existence as such does not have the same kind of relationship as a representation of a dependent existence as such except 'the other way around'. A theoretical representation instead lacks a relationship to its object that a practical representation possesses. To lack a relationship is not to possess it 'the other way around'.

The argument from akrasia and Buridan cases does not provide that account since to say that practical reason *is not* the will is not to say what it *is*. Still, intellectualism in some sense flows out of them. Since the will accounts for how we act from practical thought, practical reason is not itself tied to action. It must instead explain how practical thought has a standard of correctness that distinguishes the actions that are within my ability into the good and bad actions for me to perform. Think about that question and it might seem like there must be objects to which my judgment answers. Since that is a rough statement of the nature of theoretical reason, it is a rough statement of intellectualism. Let me explain.

Think back to the claim that my job in a practical representation is to make actual an object that I represent. Because the will has this function, I must be able to do what I decide to do, which conditions possible objects of my decision.³⁸ This restriction establishes a standard of correctness for practical thought—I can err in thought if I decide to do something that is impossible for me. It does not establish everything proper to this standard, though, since to decide to act immorally or imprudently is wrong even if the action is within my power. Vice does not become virtue if I can pull it off. It is hard, though, to get a robust standard of correctness out of the nature of the will if my job is to make actual the object that I represent. I might do that when I decide to act immorally and imprudently, and all the worse for you and me. Whence this robust standard of correctness?

There are at least three ways to go here. You might deny that practical thought has a robust standard of correctness, as old school non-cognitivists do. You might say that it has such a standard because of its form, as practical cognitivists who claim that practical reason is our kind of will do. This view must show that the idea of making actual the object of my representation precludes more than it at

38. I show in the next section that intellectualists accept this condition. I argue that it does not imply an untoward version of internalism in “Practical Possibility”.

first seems to rule out. Or you might go intellectualist and separate practical thought into components. One component, the decision, follows the model of a practical representation, and its object answers to it. The other, the judgment, follows the model of a theoretical representation and answers to its object. Practical thought has a robust standard of correctness because one component has such a standard in the way that any theoretical representation does. Yet I can act from it because of its other component.

Try it this way. Another way to contrast practical reason and theoretical reason denies the *productive* role of practical reason. It instead posits an *evaluative* role for it in contrast to the *descriptive* role of theoretical reason. An exercise of theoretical reason says that some aspect of the world is some way, an exercise of practical reason that some aspect of the world should be some way. Intellectualists endorse a version of this view, where evaluation itself is a kind of description distinguished from others because it is about reasons to act. When I judge that I should reenact the opening scene of *The Lion King* with my cat in the role of Simba, I say of an act-type that it has a specific normative property. Practical reason is thus theoretical thought about specific objects. It is the will that produces its object. According to intellectualists, it thereby is not a species of reason.

Intellectualists thus take cases of akrasia and buridan cases to show that practical reason is not the will. To say that practical reason is not the will, though, is not to say that practical reason is a species of theoretical reason. Instead of following from the separation of practical reason from the will, this intellectualist view of practical reason explains that separation. Moreover, this view seems to be able to explain the standard of correctness for practical thought. Intellectualism thereby seems to account for many of the aspects of our practical life in one fell swoop. That is its appeal.

3. *TO ACT FROM JUDGMENT*

Intellectualists break up practical thought into an exercise of practical reason and an exercise of the will.

Practical reason explains the standard of correctness for practical thought and the will explains the way that I can act from practical thought. They then must show that the way that practical reason relates to the will lets me act from practical judgment. Otherwise, what explains the standard of correctness of practical thought has no connection to action and thus cannot explain why actions are good to the extent that they meet that standard of correctness. I shall argue that their view of practical reason is incompatible with an agent acting from practical judgment. In this section, I will explain as best I can the sense of ‘from’ at work in this idea. I shall argue that intellectualists must account for it because it is a condition on the possibility of reasons for me to act.

Consider Kant’s ‘favored creature’, a kind of being whose members instinctively behave in ways that make them happy. Kant claims that “if reason should have been given ... to this favored creature, it must have served them only to contemplate the fortunate constitution of their nature ... but not to submit their faculty of desire to that weak and deceptive guidance [N]ature would have taken care that reason should not break forth into *practical use*” (Kant [1785] 4:395).³⁹ These beings can judge what would make them happy, but they cannot decide and act on the basis of this thought. They instead act from a distinct non-self-conscious capacity—their kind of will. Their normative judgments are just theoretical judgments with normative content. They cannot act *for* a reason or *take* an option *because* of a reason in the way that distinguishes rational agents like us from the other animals.

Unlike this kind of being, our practical thought has a normative relationship to action because it

39. In order to avoid pronoun ambiguity, I replace ‘it’ with ‘them’ and ‘their’ in this translation when the pronoun refers to the being and leave ‘it’ when the pronoun refers to reason. I use a plural to emphasize that the creature by nature cannot act from reason. A singular pronoun might make this creature seem like a defective member of its kind.

has a metaphysical relationship to action. Or, more accurately, the relationship between practical thought and action is at once normative and metaphysical. I exercise my capacity to act well only when I act from practical thought. I fail to exercise it well when I think that I am to ϕ but do not or if I ϕ without thinking that I am to ϕ . In order to act rationally, though, I must do more than act in a way that is accurately described by the words in the content of my practical thought. I must act *from* my practical thought, not merely *in accord with* it. After all, favored creatures act in ways that accord with their judgments, and they lack practical reason.

Although the details of the distinction between acting from something and merely acting in accord with it are difficult to articulate, this distinction is familiar. Kant uses it in order to distinguish acting from duty and merely acting in accord with it. Someone who saves me in order to achieve praise acts differently from someone who saves me because I need help. They perform different actions, and only the second acts *from* duty even though both act in accord with it. Wittgenstein likewise uses this distinction in order to distinguish following a rule and merely acting in accord with one. Someone who opens with D4 as part of a Colle opening acts differently from someone unfamiliar with chess who moves the pieces the same way because that arrangement is lovely. They perform different actions, and only the first *follows* the rules of chess even though both act in accord with those rules. My grasp of duty or the rule must in some sense be the *source* of the way that I act in order for me to act from duty or follow the rule. A practical thought has *that* kind of relationship with action. And this idea in fact structures much action theory. Just think about how ‘the same thing’ can happen when I act and when something happens to me. I might raise my hand in order to get your attention. I also might have a mind to get your attention, my hand might rise at the same speed to the same spot because of a spasm, and you

might attend to me in response to my hand rising. I act only in the first case because only then does my movement 'come from' my practical thought even though both movements accord with it.⁴⁰ And, trivially, I act well only when I act.

Intellectualists do not disagree. Although they think that the will explains the way that I act, they do not think that practical reason has *nothing* to do with action. Judging and deciding jointly constitute practical thought. In order to act well, I must judge correctly and move from judging to deciding to acting. I must act *from* what explains the standard of correctness of practical thought *through* what explains how I can act from practical thought.

Intellectualism must work this way because a reason for me to act exists only if I can act from my representation of it. In order to see this point, consider T.M. Scanlon's claim that "is a reason for' is a four-place relation, $R(p, x, c, a)$, ... that holds just in case p is a reason for a person x in situation c to do or hold a " (Scanlon [2014] 31). For example, your illness (p) is a reason for me, nurse JDF (x), on my shift in the hospital (c), to give you medicine (a). There I am, person x , a relatum in the 'is a reason for' relation that I grasp in practical judgment. This representation picks me out in a specific way. To see why, pretend I were a penguin or a doorknob. No such reason would exist. Penguins cannot represent a reason and thereby give you medicine in order to heal you. Do not even get me started on doorknobs. Similarly, even if there is gold on Jupiter, I cannot have a reason to fly there without mechanical aid. I likewise cannot have a reason to dance for the sake of sloth. And although sounds can escape from my mouth at night, I cannot have a reason to say nice things to my lover in my sleep. All such actions are in

40. Davidson and his devotees treat 'come from' as an efficient causal relationship that links metaphysically independent parts, one mental and one not, of an accidental whole. Anscombe and her acolytes treat it as a formal causal relationship that shows that a material process can be part of an exercise of a rational capacity.

one way or another impossible for human beings. The representation of me in a reason for me to act thus must pick me out as able to act from a representation of that reason. That is, it must pick me out as a human being or rational agent.⁴¹ Only then do reasons link up with my ability to act from my cognition of them in a way that explains why they are reason for *me* to *act*.

Since intellectualists treat practical reason as a species of theoretical reason that represents reasons to act, they must show that I can act from judgment. After all, since the will accounts for how we act from practical thought, action can realize a decision. As characterized, though, the will has nothing to do with reasons. Any link it might have to them must be mediated through practical reason. Otherwise, there would be no independent role for practical reason to play in an account of practical thought. It would be our will. Likewise, practical reason, as characterized, has nothing to do with action. Any link it might have to action must be mediated through the will. Otherwise, there would be no independent role for the will to play in an account of practical thought. Practical reason would be our kind of will. If intellectualism is true, I thus must be able to act from my judgment through my decision. Only then can

41. Dividing up views according to whether they think that reasons for me to act are mind-dependent is thus unhelpful. The issue is not *whether* reasons for me to act depend on my mind but *how* they depend on it. Said another way, the question is 'What is it for a reason for me to act to represent me as a rational agent?' The dispute is about the nature of our agency. One way to understand this issue is to think about Bernard Williams's famous argument for internalism about reasons to act. He starts from the premise "if there are reasons for action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action" (Williams [1980] 102). He argues from this premise for the conclusion that an agent has a reason to act in some way only if acting in some way stands in some relationship to an element of her 'subjective motivational set'. Whatever you think about this conclusion, the important point for present purposes is that it does not follow *just from the first premise*. The first premise just says that there are no reasons for someone to act that this person in principle could not recognize and thereby realize in action. It does not rule out reasons that you cannot here recognize given your motivational set or reasons that you would not be able to recognize through any procedurally rational inferences from your motivational set. The stuff about motivational sets and thus the substance of the conclusion instead comes from other premises about the nature of our agency. Since intellectualism is a partial view about our agency, it is free to specify the other premises, whatever they are, in a way that does not lead to an internalism that Williams would accept. If 'externalism' just is the denial of Williams's internalism, this view can be as externalist as you like without violating the first premise of the argument and thus denying that an agent can have a reason to act only if she can act from her representation of that reason. That is all the externalism worth wanting since to go further allows for reasons for me to act in ways that are logically, metaphysically, or physically impossible for me. Since no externalist thinks that I can have those kinds of reasons to act, they should not object to my discussion here. Their disagreement with internalists is best located elsewhere.

there be reasons to act, and only then can I possess practical reason.

4. THE POWER OF A REASON

The intellectualist view of practical reason is compatible with many views of the will. Problems with a particular compositional account might then be due to its view of the will, not the view of practical reason. I shall argue, though, that the intellectualist view of practical reason itself is incompatible with acting from judgment. I start in this section with a simple compositional account and argue that it cannot show that I can act from judgment. In the next section, I consider responses with more sophisticated views of the will. One of them seems to fix this problem. In §6, I shall argue that this response addresses only the letter of my criticism, not its spirit. In explaining why, I will show that the view of practical reason is the problem.

One way to think about whether I can act from judgment is to think about whether intellectualists can show that akrasia is a kind of failure in rational agency. If it is not a kind of breakdown between my exercise of practical reason and my exercise of the will, I am not in error when I judge that I should act one way and decide to act another way. After all, when I am akratic, my judgment and belief that I should ϕ might be non-accidentally accurate. Likewise, when I am akratic, I might act from my decision and intention to ψ . Separateness of capacities is thus not enough to show that akrasia is possible, at least so long as akrasia is a kind of failure in rational agency.

In fact, *mere* separateness entails that practical reason and the will have nothing to do with each other. If practical reason is a capacity to apply concepts, it answers to its object. If the will is a capacity to realize concepts, its objects answer to it. If there is nothing else to say, my decision does not answer to

my judgment, and akrasia is not a kind of failure. Moreover, because I act from my decision, I cannot act from my judgment. My action merely accords or discords with judgment.

No intellectualist thinks that the relationship between judgment and decision does not partly determine whether I decide well. They must explain the relationship between practical reason and the will such that this claim is true and show that it allows me to act from judgment. Might they invoke a reason in order to explain why?⁴² After all, the issue about deciding well is a normative matter. If I act well only if I act from my judgment, maybe I base my decision on my judgment because I have reason to decide in that way. This response treats the question about the relationship between practical reason and the will as a normative question, not a metaphysical question. It is as if I ask ‘Should I base my decision on my judgment?’, and the answer is ‘Yes, because I have reason to decide in that way’.

On reflection, this response might seem obvious. After all, the object that determines whether a practical judgment is correct is a reason for me to act. If to act is to realize the object of a decision, a reason for me to act is a reason for me to realize the object of a particular decision. I can do that only if I make that decision. The object of a practical judgment might then seem to show that I should decide on the basis of my judgment. To represent one and still ask about how to decide might then seem to be asking for “a reason to ... do ... what [I judge that I have] reason to ... do” (Parfit [2011] II.420). This is close to a “nonsensical” question that “ask[s] for a reason” to do as my reasons say (Scanlon [2014] 62).

In fact, though, a reason cannot play this explanatory role. According to intellectualism, to recognize a reason is to represent it in an exercise of practical reason. It thus determines whether a

42. Or, what is the same thing, a brute normative fact that I should decide to act as I judge that I should act. Depending on how you interpret it, this response is materially equivalent to either the reason response in this section or the view in the next about the principle of the will.

particular practical judgment is correct. In this way, there cannot be *reasons to decide* in the sense of reasons that are the grounds for my decision. If there were, the will would play its role in rational agency and the role of practical reason as intellectualists think of it. In other words, practical reason would be the will. The separation of practical reason and the will thus precludes the possibility of reasons to decide. However, if nothing shows that I decide well only if I base my decision on my judgment, invoking a reason that determines that a particular practical judgment is correct does not help. In order for this reason to matter for my decision, my will must answer to the judgment that represents that reason. My will only answers to that judgment, though, if it answers to practical reason in general. Since my will as yet answers to nothing, this reason and my judgment about it are irrelevant to my decision.

Try this analogy. Say you tell me to ϕ , I ask why I should listen to you, and you reply 'you should do as I tell you', not as a joke or a threat but as an explanation of your authority over me. The problem is obvious. Declaring your authority does not establish it. The authority of that declaration is in question because whether you are boss is in question. That declaration has authority over me only if you in general have authority over me. In the same way, intellectualists cannot show that my decision answers to my judgment by invoking a reason to decide on the basis of my judgment. The reason determines which practical judgment is correct. My judgment has authority over my decision, though, only if practical reason has authority over the will. That is what is in question. The will does not answer to practical reason if it is one capacity whose objects answer to it, practical reason is another that answers to its objects, and they have nothing to do with each other. A reason and a correct judgment that my exercises of these capacities should have something to do with each other cannot solve this problem.

This fact also shows a deeper problem with invoking a reason here. Reasons for me to act exist

only if I can act from judgment. Practical reason cannot explain this ability since it is a species of theoretical reason that answers to its object. Without an explanation, though, no reasons for me to act exist, including a reason to make and realize a particular decision. Since intellectualists separate practical reason from the will, they need to show that these capacities relate to each other such that whether I exercise my will well at least partially depends on whether my decision answers to my judgment. Otherwise, I cannot act from my judgment, and I cannot have any reasons to act. To invoke a reason in response to my criticism is thus a 'nonsensical answer' because it invokes a particular reason in order to explain a condition on the possibility of reasons as such.

5. THE PRINCIPLE OF THE WILL

The separateness of capacities threatens to make acting from judgment impossible. However, this problem might seem to be a quirk of the previous compositional account of practical thought. In particular, it might seem to follow from an optional view of the will. Mere separateness of capacities says that practical reason answers to its objects and the objects of the will answer to it. This stark view isolates practical reason and the will in the sense that they function independently of each other. Why not alter the views of these capacities, though, so that the function of one depends on the other? In particular, why not change the view of the will? After all, to say that the objects of the will answer to it is not to say that the will does not answer to *anything*. It is only to say that it does not answer to its objects.

Intellectualists might first try to get the will to answer to practical reason by letting reasons once again play the crucial explanatory role. Instead of invoking a reason to decide to act as I judge that I should act, though, what if my decision answers to my judgment because reasons to act determine the

standard of correctness of both the judgment and decision? Say that the principle of practical reason is to judge that I should do what I in fact should do and the principle of the will is to decide to do what I in fact should do. This account might seem to explain the irrationality in akrasia. If reasons to act determine both which judgment is correct and which decision is correct, I must be either judging incorrectly, deciding incorrectly, or both anytime that I judge that I should do something but decide to do something else. How, though, does this account show that practical reason has authority over the will? For one thing, the reasons to act determine which decision is correct independent of any exercise of practical reason. That is why I should not decide to act as I judge that I should act when my judgment is incorrect. Hence, the problem with akrasia is not that my judgment conflicts with my decision. That conflict is just evidence that there is an issue with the judgment or decision themselves, not in relation to each other. On the same grounds, although I should decide to act as I judge that I should act when I judge correctly, I should not decide to act in that way *because* I so judge. I should instead decide to act in that way because of the reasons regardless of how I judge. Far from showing that practical reason has authority over the will, this view instead shows that my exercise of practical reason is irrelevant to my exercise of the will.

For another thing, on this account, practical reason has no role to play in an account of our agency. After all, the will is now a capacity to recognize reasons to act and act in ways that realize its representation of those reasons. It thereby plays both the role of practical reason and the role of the will in an account of practical thought. Instead of being two capacities, this view of the will makes practical reason our kind of will.

Intellectualists might try to block this argument by saying that although reasons to act determine

both which judgment is correct and which decision is correct, I can appreciate reasons to act only in an exercise of practical reason. Perhaps then my will must respond to my judgment because it is the only way it has a shot of tracking the reasons to act. Yet this limitation of the will is unmotivated. After all, if I can recognize and respond to the verdict of practical reason in an exercise of my will, why can I not recognize and respond to the reasons in that exercise? What distinguishes those reasons from my judgment? In effect, intellectualists face a dilemma here. On the one hand, say that I can recognize and respond to reasons to act in an exercise of the will. Practical reason then has no independent role to play in an account of practical thought. The will plays its role and the role of practical reason, which is to say that practical reason is our kind of will. On the other hand, say that I cannot recognize and respond to reasons to act in my exercise of the will. Why? I can recognize *something* in that exercise since intellectualists think that I can recognize my judgments about what I should do. Why can I recognize them but not reasons to act? It must be something special about reasons to act that distinguishes them from judgments. But what? Whatever intellectualists say here will also apply to practical reason. After all, if the principle of both capacities say that I am to exercise them in a way guided by reasons to act, their principles cannot explain why I can recognize reasons to act in the exercise of one but not the other. Whatever explains why I can recognize them in an exercise of practical reason will say that I can recognize them in an exercise of the will. Whatever explains why I cannot recognize them in an exercise of the will will say that I cannot recognize them in an exercise of practical reason. If intellectualists go with this view of the will, then, they must either say that practical reason is our kind of the will or say that human beings lack a capacity to represent reasons to act and thereby say that we lack practical reason. Those options are no good.

Intellectualists instead should say that although practical reason answers to its object and the objects of the will answer to it, the will also answers to practical reason because its principle is ‘decide to act and act as I judge that I should act’. On this account, the “will is the [capacity] to choose and perform intentional actions. We exercise our will when we endorse the verdict of reason that we must perform an action” (Raz [1998] 48). In other words, “the function of [practical] reason ... is to assess various reasons for action and arrive at conclusions about which actions there is sufficient reason to perform. The function of the will is then to determine which of these eligible actions we will in fact do” (Scanlon [2004] 231-2).⁴³ As we might put the idea, the principle of my will is to decide to do a determinate version of what I judge that I should do or, in Buridan cases, a determinate version of one of the act-types that I judge are on a par. Hence, my “actions ... are guided by and responsive to” my judgments because of something that is not a “merely classificatory norm” but is instead “grasped and applied in a way that directly gives rise to action” (Wallace [1999] 44). And when the will “fails to be guided by judgement, it fails to operate in its executive capacity—it fails to operate *as a will*” (Watson [2003] 136). Akrasia is then a breakdown in our agency because in it I exercise my will badly—I exercise it in a way that conflicts with its principle.⁴⁴

43. Scanlon is describing Raz’s view. It is unclear whether he himself endorses this view at that time. He does not endorse it in more recent work. Still, it clearly states a common view that responds to my previous criticism of intellectualism.

44. Let me discharge a debt. Christine Korsgaard criticizes *dogmatic rationalism*, which says that all normativity is the normativity of substantive reasons, or at least is based on it (Korsgaard [1997] 52-4; [2003] 315-7; [2008_a] 24-30; [2009] 64-7). Her arguments are complicated, but she concludes that the normativity of substantive reasons depends on a kind of functional normativity wherein, in my terminology, a capacity establishes a normative standard for its exercises. I agree with this argument, but it does not undermine intellectualism. After all, theoretical reason is a capacity. Intellectualists can say that judgments answer to their objects because of the function of theoretical reason, and practical reason is a species of theoretical reason distinguished by its objects. This seems to let them capture everything that matters to dogmatic rationalists, or at least everything that should matter. Korsgaard and I disagree with this view, but nothing about the functional normativity of capacities refutes it.

Korsgaard recognizes this possibility when she claims that dogmatic rationalists must think of practical judgments as knowledge that leaves open the question about what to do with it. However, she does not consider whether they might agree with this claim because they think that the will is a separate capacity with its own function by which an agent can close this

I shall argue in the next section that this response only addresses the letter of my criticism, not its spirit. It does not show that I can act *from* judgment in the sense needed in order to act for a reason. It in fact reveals that focusing on views of the will distracts from the real issue. This argument is dense and difficult. I will use analogies in order to get the idea in view before giving the metaphysical argument. Let me first, though, explain the conclusion for which I shall eventually argue. Please bear with me.

Consider the following claim by Kieran Setiya. Setiya agrees that “it would be absurd to regard the connection between these two kinds of commitment as wholly contingent: while it is possible to decide against one’s better judgement in a given case, one cannot do so all the time” (Setiya [2005] 787). Most intellectualists try to explain why the link between exercises of practical reason and exercises of the will is not ‘wholly contingent’ with an account of the will that shows that its exercise is per se such as to respond to the exercise of practical reason. Setiya suggests that they should instead approach things the other way around and say that “nothing could count as practical judgement if it had no tendency to issue in choice” (ibid.). Talk about the ‘tendency’ for an exercise of practical reason to issue in an exercise of the will is vague, but Setiya clearly means that an account of practical reason should show that it is per se such as to determine the will. After all, if the will is per se such as to respond to exercises of practical reason, an exercise of practical reason tends to issue in an exercise of the will. This tendency, though, is an *external* tendency of practical reason—a tendency that exists not because of the nature of practical reason but because of the nature of the will. Setiya instead thinks that intellectualists should posit an internal tendency of practical reason—a tendency that exists because of the nature of practical

question. Again, while she and I disagree with this view, the functional normativity of capacities cannot undermine it. The intellectualist view that I develop in this section thus avoids the letter of her objection. It does not avoid its spirit, though, which I hope my criticism embodies and emboldens.

reason.⁴⁵ In my terminology, he thinks that they should claim that the principle of practical reason links its exercise with the exercise of the will rather than or at least in addition to saying that the principle of the will links its exercise with the exercise of practical reason. The nature of practical reason should explain why it is such as to determine the will.

My argument in this next section is in effect that intellectualists *must* take up this suggestion but that their view of practical reason rules it out. I will argue that they must try to show that practical reason is per se such as to determine the will because otherwise they commit themselves to the possibility of a kind of being who possesses and exercises practical reason perfectly even though she lacks a will. Such a being is metaphysically impossible, though, and whatever capacity she might have is not practical reason, even on the terms in which intellectualism understand practical thought. However, intellectualists cannot say that practical reason is per se such as to determine the will. Practical reason on their view has the formal features of any species of theoretical reason, and species of theoretical reason are not per se such as to determine the will. Hence, intellectualism is untenable because of its view of practical reason, regardless of the view of the will that any of them endorse. So I shall argue.

6. The Isolation of Practical Reason

Start with this analogy. I live in the United States and am subject to its laws. Now imagine if its only law is ‘follow the laws of Luxembourg’. To commit a crime, then, is to act in a way that discords with the laws of Luxembourg. I am subject to the laws of America, though, not of Luxembourg. Although their laws

45. Because Setiya thinks that intentions are ‘desire-like beliefs’ in *Reasons without Rationalism*, he might seem like someone who thinks that practical reason is a species of theoretical reason but does not separate practical reason from the will. However, these quotes and his insisting in that book on the possibility of clear-eyed akrasia wherein I intend to act in some way while knowing that so acting is not good show that he does not unite practical reason and the will. The view of intentions is instead a view about the nature of the will as a capacity distinct from practical reason. Setiya might be an intellectualist about practical reason as well, but his view of the will is neither here nor there for his view of that capacity.

determine the content of ours, they lack authority over me. They merely inform something else that has authority over me, and even then only because that something else says so. That is the laws establish different jurisdictions. In the same way, on this account, my decision is subject to the principle of the will. Although that principle gets determinate content in particular situations from my judgment, practical reason lacks authority over the will. My judgment is relevant to my decision only because it informs something else that has authority over my decision, and even then only because that something else says so. When the principle of my will is 'decide to act as I judge that I should', I err by deciding in a way that discords with my judgment. I err with respect to the principle of my will, though, not my judgment, just as I violate our law, not Luxembourg's. Although my judgment determines the answer to my volitional question, it just informs my will, and I act from my decision alone. Practical reason is still isolated from the will in the way that matters. I still cannot act from judgment.

Although intellectualists need my judgment to determine the way for me to will well, then, not everything that meets that description lets me act from judgment. To get a sense of the issue, think about relationships in which one person has authority over another. My boss has authority over me with respect to the tasks of my job. I am subordinate to her with respect to them. My teacher has authority over me with respect to the tasks of the course. I am subordinate to her with respect to them. In each case, our roles are essentially interdependent. Each role is intelligible only in relation to the other.

Now say that we do not stand in a relationship of this kind, but I set out to decide on the basis of your thoughts. No authority relationship holds. A sycophant is no subordinate. That is why criticism that is appropriate in those relationships is inappropriate in this case. If you are my boss and my job requirements are unclear, I can criticize you for it. If I am not fulfilling my tasks, you can criticize me for

it. This criticism is appropriate because we embody roles defined by the relationship between them. If I am a sycophant, I cannot criticize you because your thoughts are too unclear for me to base my decisions on them. Nor can you criticize me if I fail to act on the basis of your thoughts. That kind of answerability does not hold because the roles are not essentially interdependent. Even though we cannot understand my role in this relationship without relating me to you, we can understand yours without relating you to me. You then do not fail in your role in this relationship if your thoughts are unfit to be taken up by me in decision and action. If I take them up in these ways, I make use of them for my own purposes. I then cannot act from your thought even if I base my decision on your thought.

The problem with the current compositional account is that it treats the will like a sycophant. It is as if practical reason is the name of one person who judges, the will of another who decides on the basis of those judgments and acts. Just as your role in our sycophantic relationship is fully described without mentioning me, so practical reason is fully described without mentioning the will. It does not by nature have a relationship to the will because it only answers to its object. The only relationship that it has to the will and action is then accidental to its nature. It thereby does not fail if its judgments are unfit to be taken up by the will. If the will takes them up, it makes use of them for its own purposes. I then cannot act from judgment even if I base my decision on my judgment and even if to decide well is to decide on the basis of my judgment. The source of the issue is not the nature of the will. It is the nature of practical reason. Separate practical reason from the will and it cannot have an essential relationship to the will. It can only have an accidental relationship in the same way that you only have an accidental relationship to me when I am your sycophant.

The problem is not that intellectualists cannot say that the will has an essential link to practical

reason. The problem is that they cannot say that practical reason has an essential link to the will. To appeal to the analogy with the sycophant one last time, the problem is that practical is not per se such as to determine the will in the same way that the celebrity is not per se such as to determine the will of the sycophant. Just as the proper description of the celebrity in her role as a celebrity does not reference the sycophant, so the proper description of the nature of practical reason does not reference the will. Yet then just as the celebrity can exist as a celebrity without the sycophant, so an agent can by nature possess practical reason as intellectualists think of it while by nature lacking a will. A being without a will, though, cannot act, and a being who cannot act cannot possess practical reason. The intellectualist account of practical reason thus implies that a metaphysically impossible being is possible. The failure to explain this impossibility indicts intellectualism.

Thus far, my argument in this section is analogical. I shall now give a metaphysical argument to the same conclusion for which the analogies have prepared the way. In order for me to be able to act from judgment, practical reason must have authority over the will. If it is a species of theoretical reason, either one of its distinct features explains why only it of all species has authority over the will or theoretical reason in general and thus all of its species has authority over the will. Neither option works. Start with the first. Species of theoretical reason differ from each other because of their objects and the corresponding content of their representations of those objects. Mathematical reason is about mathematical objects and the content of mathematical judgments reflects it, chemical reason is about chemical objects and the content of chemical judgments reflects it, and all that jazz. Can one of these features show that only practical reason of all species has authority over the will? The object cannot. Since exercises of theoretical reason answer to their objects, those objects just determine which

judgments are correct. They do not determine whether I must do something else after I judge and believe. The content of practical judgments is then the only thing that might explain the authority of practical reason if that authority is unique among species of theoretical reason. That distinctness, though, just reflects the fact that species of theoretical reason differ because of their objects. It does not distinguish the way that I relate to practical objects and practical judgments from the way that I relate to objects of other species and my representations of them. It thereby does not show that I must do something else after I judge and believe when my judgment and belief are about practical objects. The nature of theoretical reason determines those things. Practical reason then has authority over the will only if theoretical reason in general has it.

But theoretical reason lacks this authority. Take my judgment that a ball is in the room. This judgment licenses different inferences than the judgment that a cow is in the closet licenses. I can infer that members of a ball-making species or someone who came into contact with them was in these parts. (Probably it was humans.) I can infer that there is at least one object in the room. The first judgment licenses the others because its content, or the object that it represents, constitutes grounds for them according to the principle of the capacity that I exercise in those other judgments. Any exercise of theoretical reason answers to its object. The object is an independent existence that I represent as such. And that is it. That exercise of theoretical reason is done and dusted. Anything else I do is a discrete exercise of one of my capacities that answers to the principle of that capacity, not to the previous judgment, though it may be informed by that judgment if the principle says so. If practical reason is a species of theoretical reason, to exercise it is to represent an independent existence as such. And that is it. Anything else I do is a discrete exercise of one of my capacities that answers to the principle of that

capacity, not to the previous judgment, though it may be informed by that judgment if the principle says so. Practical reason thus lacks authority over the will. The nature of theoretical reason precludes it.

One way to understand this idea is to think about the content of a practical judgment. I keep saying that I judge that 'I should ϕ '. However, intellectualists tend to say that I judge that ' p is a reason for X to A in circumstances C ' or that ' X has most/sufficient/decisive reason to A in circumstances C '. If we call any judgment about reasons to act a 'normative judgment' and call only a (present tense) judgment about *my* reasons to act a 'practical judgment', intellectualists think that practical reason is a capacity for normative judgment. Practical judgments are just normative judgments that happen to be about me—that have a particular value for x . According to intellectualism, they are unique among normative judgments in the same way that normative judgments are unique among descriptive judgments—because of their content, not because of the way that I relate to the object or my representation of it.

It might seem like a virtue that intellectualism can explain the unity of normative judgments. What seems like virtue, though, is vice. I do not relate to my representations of yellow objects differently than I relate to my representations of blue objects. According to intellectualism, I likewise do not relate to my representations of reasons *for me* to act differently than I relate to my representations of reasons *for you* to act. Whether any of my normative judgments refer to me is an accident just like it is an accident whether any of my judgments about spatial relations refer to me. It is not that it is unlikely that many of these judgments will refer to me. Given my finitude, it would be shocking were it otherwise. It is not part of the nature of reason, though, that any of my normative judgments or judgments about spatial relations refer to me. When one does, it refers to me in the way that another refers to you. Anyone can

represent these objects, and everyone relates to the objects and to their representations of those objects in the same way. I understand in a thought about a spatial relationship involving me that you can represent what I represent and that our thought relates to it in the same way. According to intellectualism, I likewise understand in a thought about a reason for me, for you, or for someone else to act that you can represent what I represent and that our thought relates to it.

The only judgments that have authority over action, though, unite a first-person representation with an action I take to be open to me. Here are some that lack this authority.

- (5) Judgments *that I should have ϕ -ed* lack authority over action. To not act on the judgment that I should not have taken up self-directed study of the kazoo when I was 5 is not irrational. I might be irrational for not making other judgments on this basis, like the judgment that I should make amends to others—my parents, no doubt, probably the neighbors, and maybe even poor Sadie dog—if I wronged them by acting in this way. I am practically irrational if I do not act on that judgment. It, though, is a first-person judgment about an action I take to be open to me. Absent the possibility of time-travel, anyway, I am not practically irrational for not acting on a judgment about the past.
- (6) Judgments *that you should ϕ* lack authority over action. To not act on my judgment that you should comment on my paper is not irrational. I might be irrational for not making other judgments on this basis, like the judgment that I should ask you to comment on the paper or that I should stop distracting you and let you get to it. I am practically irrational for not acting on those judgments. Those judgments, though, are first-person judgments about actions I take to be open to me. I am not practically irrational for not acting on a judgment that is not about me.
- (7) Judgments *that, in counterfactual circumstances, I should ϕ* lack authority over action. To not act on my judgment that I should join the circus were I able to fly without mechanical aid is not irrational. Absent the truth of modal realism and the possibility of world-travel, anyway, I am not practically irrational for not acting on a judgment about non-actual circumstances. I am not irrational in this way even if it turns out that these circumstances obtain. Maybe I can fly if I think lovely wonderful thoughts while you sprinkle a bit of fairy dust on me, but I have no knowledge of these powers of the dust. I might be irrational for not knowing about them. That is theoretical irrationality, though, not practical irrationality. I might come to know about them and infer that I should join the circus. I am practically irrational if I do not act on that judgment, but it is a first-

person judgment about an action I take to be open to me. I am not practically irrational for not acting on a judgment that includes in its content a claim that the action is inappropriate in the circumstances as I understand them.

- (8) Judgments *that that person should ϕ* that refer to me only demonstratively or under a definite description lack authority over action. To not act on the judgment that the mysterious man in the loud shirt should wipe his face is not irrational even if I am that man in the mirror. I might not know that I am him and he is me. I might be irrational for not knowing that I am that man. That is theoretical irrationality, though, not practical irrationality. I might come to learn that I am that man and judge that I should wipe my face. I am practically irrational if I do not act on that judgment, but it is a first-person judgment about an action I take to be open to me. I am not practically irrational for not acting on a judgment that I do not take to be about me.

Although these judgments might be criticizable, I am not irrational in not acting on them. How could I act from my sense of why I act when I take my representation to be irrelevant to what I should do here and now? I could not! In this way, only first-person judgments about actions I take to be open to me have authority over action. Only practical judgments have authority over action.

However, intellectualists cannot explain this unique feature of practical judgments. They must say that I relate to every reason to act and my representation of it in the same way no matter if it is about you or me, the past or the present, a merely possible world or the actual world, or any other difference that distinguishes one from another. I relate to reasons to act and my representations of them just as I relate to any object of theoretical reason and my representation of it. They are all independent existences that I represent as such, and you relate to them in the same way. That is what it is for practical reason to be a species of theoretical reason. To say otherwise is like saying that my judgment that the ball is yellow is subject to different rules of inference than my judgment that the ball is blue.

My representations of most reasons to act and of all other objects of theoretical reason, though, lack authority over my will. Hence, my representations of reasons for me to act lack that authority. Just

as our judgments that the other should ϕ do not have authority over the way that either of us act, so your judgment that you should ϕ and my judgment that I should ϕ do not have authority over the way that either of us act. We cannot act from those representations anymore than I can act from my representation of a reason for you to act, your representation of a reason for me to act, or either of our representations of a pineapple. The source of the issue is the form of representation. Nothing about the content or, for that matter, the details of the nature of another capacity can change it. Non-accidentally representing an independent existence accurately by nature has nothing to do with deciding and acting. Hence, I cannot act from judgment. Practical reason is isolated from the will and action in us just like it is isolated in Kant's favored creature.

This consequence follows solely from the nature of practical reason on an intellectualist account of it. The current account posits an asymmetric metaphysical dependence between practical reason and the will. According to it, any being who by nature possesses this kind of will by nature possesses practical reason. According to the intellectualist view of practical reason, though, a being who by nature possesses it need not possess a will at all, let alone a particular kind of will. After all, if one capacity has an accidental relationship to another, a kind of being can by nature possess the first while by nature lacking the second. On this account, then, a being can possess the same capacity of practical reason that we possess while by nature lacking a will and the ability to act. Call this kind of being the 'unfavored creature'. They are like a bit of water who tells of all she would do if she were to 'decide' to no longer willingly sit idle in this puddle, some leaves who narrate their descent through the air, or a piece of paper that 'decides' to go as the wind blows (Schopenhauer [1839] 42; Anscombe [1957] §3; Wittgenstein [1939] 434). They cannot act from judgment since they cannot act at all. According to intellectualism,

we possess the same capacity of practical reason that these kinds of beings possess. After all, practical reason only has an accidental relationship to the will and action. Nothing, then, differentiates their judgments *as such* from our judgments *as such*. They can make the same judgments as us with the same correctness conditions and the same intrinsic features. Hence, we cannot act from our judgment any more than any of them can act from theirs even if those judgments inform our decisions. Informing decisions is an accidental property of those judgments that has no place in an account of their nature. Practical reason is thus isolated from the will and action if it is a species of theoretical reason no matter what view of the will a particular compositional account of practical thought posits.

In effect, then intellectualism faces the following problem. Since any being with practical reason is a rational agent and since intellectualism implies that a being can have practical reason without having a will and hence without being able to act, intellectualism implies that there are rational agents who cannot act. That is to miss the *agency* in rational agency. The only way for intellectualists to respond is to deny that a being can possess practical reason as they think of it without possessing a will.

Intellectualists might try to argue for that claim in the following way. Practical reason is about reasons to act. To act is to realize a decision, and hence the will has an essential connection to action. A reason to act is then a reason to realize a decision, and it can exist only if the being who is to act can decide and hence has a will. Hence, practical reason might seem to have an essential connection to the will through reasons to act, not because these reasons determine which judgment and decision are correct as in the earlier view but instead because reasons for me to act imply that I can act and thus have a will.

However, this argument is fallacious. Although it establishes that I cannot have practical reason

as intellectualists think of it unless *someone* has a will, it does not show that I cannot have practical reason unless *I* have a will. Again, the source of the problem is the view that practical reason is a species of theoretical reason distinguished from other species by its object. Theoretical reason does not restrict the possible contents of the judgments that I can make to judgments that can truly apply to myself. Just as I cannot make theoretical judgments about gills unless some beings can have gills, I cannot make theoretical judgments about reasons to act unless some beings can have some and hence have a will.⁴⁶ Just as I can make theoretical judgments about gills without having some myself, though, I can make theoretical judgments about reasons to act without having some myself. So intellectualists must say. Hence, a being can have practical reason as intellectualists think of it while by nature lacking a will and thus lacking reasons to act. They cannot act from their practical judgments since they cannot act at all. Since their judgments are intrinsically identical to our judgments, we cannot either.

The intellectualist capacity for theoretical judgments about reasons to act is thus no capacity of practical reason, even on their own terms. If I cannot act from judgment, I cannot act *for* a reason or *take* an option *because* of a reason in the way that distinguishes rational agents like us from the other animals. Intellectualists thus must deny that we are rational agents who have reasons to act in the sense that those italicized phrases mean to describe. Without reasons to act, though, there cannot be a species of theoretical reason about them. Without an essential connection to what explains how we can act from

46. I here restrict attention to 'positive' judgments that say that someone has a reason to act in some way or that someone should act in some way. I make no claims about whether intellectualism implies that a being can have practical reason while no being has a will because she can judge that no one has any reason to act. Intellectualists can deny such possibility if they can deny that such a being could possess the concept of a reason to act. While I doubt that this denial actually fits with intellectualism, I do not need it for my argument here. Since intellectualists treat practical reason as a species of theoretical reason, the concept of a reason to act must be within our reach in the same way that the concept of a gill is within our reach. Each of them picks out a kind of object that we represent in the same way. Hence, even if a being cannot make practical judgments, positive or negative, if no one has a will, she can make positive and negative practical judgments if anyone has a will. That is all that I need for my argument.

practical thought, then, practical reason is lost. Without practical reason, we cannot act because of our sense of why we act since we cannot have a sense of why we act. We cannot act from practical thought, and the idea that we act self-consciously is lost. Try living with that.

7. REJECTING THE LEGEND

Intellectualism separates practical reason from the will. To deny it is to say that practical reason is our kind of will. This view might seem hard to maintain. For one thing, intellectualism seems to fall out of the description of Buridan cases and akrasia. Intellectualism might then seem compulsory, my criticism be damned. For another thing, intellectualism seems to be able to account for the robust standard of correctness for practical thought. What other account is available? Finally, what about the relationship between practical judgments and other normative judgments that intellectualism unifies? Without something to say about these issues, you, dear reader, might insist that there must be an error in my criticism. I cannot articulate and defend an alternative view here. Let me freestyle a bit, though, and I will show that the intellectualist account of akrasia and Buridan turns on an optional untheorized metaphysical assumption. Likewise, I can explain the possibility of a robust standard of correctness for decisions, and I can say something about the relationship between practical judgments and other normative judgments. It is not everything, but it is a start.

Start with akrasia. Intellectualists infer from the possibility of a breakdown in rational agency to the conclusion that practical reason and the will are separate capacities. While this inference might seem trivial, it actually turns on a untheorized metaphysical assumption. There are at least two ways to think about things coming apart. On one view, it shows that a whole is an accidental unity of distinct things

that do not by nature belong together. Breakdown shows things as they are, which we might miss when we look at the composites that they can form. If the oxygen can separate from the hydrogen, say, water is not fundamental. It is instead composed of independent parts that have no intrinsic relationship to each other. Intellectualists view breakdown on this model. The isolation of practical reason follows. On the other view, the fact that things can come apart does not always show that they are distinct things that happen to come together. Some wholes are prior to their parts, and the parts are per se parts of the whole even if they come apart in pathological cases. Although I might lose an arm, say, I do not consist of at least two independent parts, that arm and the rest of my body, that do not by nature belong together, though none of us, especially not me, might notice it until things break down. The way to think about akrasia is on this model.

More must said, but this view has an obvious advantage over the other. Akrasia, if it is anything, is a failure in the exercise of human agency. It is our agency *breaking down*. Separate practical reason from the will and you have a hard time explaining this fact. What is wrong when two things that do not by nature go together do not go together in a particular case? On this model, finding a judgment without a decision should be on a par with finding an atom of oxygen without two atoms of hydrogen. The view that unites practical reason with the will does not have this problem. If the aspects of a practical thought that explain its standard of correctness and how I can act from it by nature are two aspects of one thing, a situation where they come apart is a per se defective practical thought. It is like finding a human arm without the rest of a human body—disturbing, unnatural, something gone wrong.

Uniting these aspects of practical thought also solves the issue with Buridan cases. If you think that practical reason is a species of theoretical reason, these cases are pressing because practical reason

cannot distinguish between actions that are normatively on a par. An abyss opens up between reason and action that reason cannot cross. If practical reason is our kind of will, it differs from theoretical reason on formal grounds. It is in part a capacity to realize concepts. In practical thought, I move from a general representation of what to do and why to do it through more determinate representations of how to do it given my circumstances and skills. At some point in this thought, the normative differences between different ways to realize this action run out. Still, the work of practical reason is not just to represent an action, or even to represent the best available action in my circumstances. The work is in part to act—to realize a representation. So long as a way to act realizes the object of my representation, it meets this part of its principle even if other ways also meet it. Hence, acting in a way that I recognize is normatively on a par with other ways of acting open to me can be an act of reason or part of an exercise of practical reason if practical reason is our kind of will. There is no need to posit another capacity.

Of course, although practical reason is in part a capacity to realize concepts, I can be mistaken if I represent the wrong action from among the actions within my abilities. Otherwise, we are not subject to moral standards and prudential standards in action. I cannot here say anything substantial about the way to account for a robust standard of correctness in a non-intellectualist fashion, though I make a start of it in “The Substance of Constitutivism”. The general shape of the answer, though, is present even in intellectualism. After all, intellectualists think that practical reason is a species of theoretical reason, and theoretical reason is a capacity to non-accidentally represent distinct existences accurately. The fact about the *nature* of the capacity establishes standards of correctness for its exercises. Because theoretical reason is a capacity to non-accidentally represent distinct existences accurately, I exercise it well when I do so with respect to a particular object and badly when I accidentally represent an object accurately or

represent the object inaccurately. This explanation has the form of what I elsewhere call *constitutivism* that says that a single principle describes the nature of a capacity and is thereby normative for its exercises. It is a particular kind of relationship between something general and the particulars that fall under it that I claim holds for all things that have internal normative standards. The nature of practical reason establishes an internal standard of correctness for its exercises, then, just as much as the nature of theoretical reason establishes an internal standard of correctness for its exercises. Both of them are capacities. The difference between them is the content of the principles that describes the nature of these capacities and thereby establishes different standards for their exercises.

The issue with respect to practical reason, then, is not *how* it can have a standard of correctness. It is about the content of that standard. The idea of the will already brings with it some content. In order for me to exercise my will well, I must do the thing that I decide to do. This fact places constraints on my decisions. I cannot do the thing that I decide to do if it is impossible for me. It is then an internal standard of adequacy on a practical thought that the action that I represent be possible for me. I thus can have a reason to act in some way only if I can act in that way from my representation of that reason. How exactly to understand this condition is another issue. The crucial question is about how to understand the representation of myself in a reason for *me* to act. Everyone agrees that this representation picks me out as a rational agent. Different accounts of rational agency, and in particular different accounts of the relationship between rational agency and human rational agency, will result in different substantive standard of correctness. The possibility of morality or prudence depends on the possibility of an appropriately substantive account of our agency that starts from the fact that practical reason is our kind of will. I cannot offer that account here and so cannot articulate the content of the standards of

correctness for practical thought. However, the question about whether there can be such a standard is answered. Practical reason is our kind of will. Its nature establishes standards for its exercises just like the nature of any capacity—including the capacities that intellectualists invoke—establishes internal normative standards for its exercises.

Absent the account of our agency, I cannot say what anyone should do or why. However, I can explain what it is for me to say what someone should do in a way that explains the relationship between practical judgments and other normative judgments. Because I am self-conscious, I can reflect on my exercises of practical reason and grasp its standard of correctness, at least so long as I have sufficiently developed this capacity through training, education, practice, reflection, and everything else that goes into a life. I can recognize that this capacity can have a manifold of bearers who exercise it a manifold of times. I can think about how the standard of correctness bears on different situations and agents without being in those situations or being those agents. My judgments about these things will be theoretical judgments with normative content about what follows from the principle of practical reason for people who fall under it.

In order to make these theoretical judgments, I must have the concept of a practical judgment or a reason to act. I cannot get these concepts through theoretical reason, though, because of the first-person nature of practical judgment. I can only get them through reflection on my exercises of practical reason. The ability to form these theoretical judgments thus depends on the ability to exercise practical reason. The concept of a reason to act is thereby an essentially first-personal concept. Only beings who can act for reasons can possess it, and we come to possess it in the course of developing our capacity to

act.⁴⁷

Once I possess the concept of a reason to act, though, I can use it in theoretical representations about others or myself. In this thought, I think of that person in some specified circumstances and say that this person should act in some way. I in effect say that this person should realize a specific practical thought or that making a particular decision and acting accordingly meets the internal normative standards for exercises of their capacity to act. To act is, after all, to realize a practical thought, and different actions are distinguished by which thought they realize. My theoretical judgment is correct or meets its internal normative standard if it non-accidentally represents its object accurately. That object is a practical thought, which is correct if it meets its own internal normative standard. Hence, the theoretical thought is correct in its way only if the practical thought in question is correct in its way. Self-knowledge of my self-conscious capacity to act and my first-person grasp of reasons for me to act is thereby the basis for all of my cognition about reasons to act, whether for me or others, whether in actual circumstances or not, in the past, present, and future. But not all cognition of reasons to act is practical cognition.⁴⁸ Practical thought is productive thought, but theoretical thought about practical thought is

47. Kant's favored creature, my unfavored creature, Schopenhauer's water, Anscombe's leaves, and Wittgenstein's paper are then metaphysically impossible. They cannot have thoughts about reasons. Kant does not disagree. He presents his creature in the service of a reductio of a view of practical reason. He argues that if the function of practical reason were to secure happiness, then, given some assumptions about when beings possess capacities, no being would possess practical reason. Since we possess practical reason, the function of practical reason is not to secure our happiness. The favored creature is used to illustrate what a being designed to secure happiness who also had reason would be like *if such a being were possible*. This argument does not entail the metaphysical possibility of such a being anymore than my criticism of intellectualism entails the metaphysical possibility of the unfavored creature. Likewise for Anscombe, if I understand *Intention*.

Schopenhauer and Wittgenstein, the latter no doubt in the voice of an interlocutor, disagree. They offer their images in order to debunk the idea that we act self-consciously. According to them, we are like the water and paper, though we do not realize it. If the position I am sketching works, this kind of scepticism about practical reason is untenable. The water and paper could possess the concepts needed to think these thoughts only if they had practical reason, but if they did they would be able to act self-consciously. I do not doubt that Wittgenstein, in his own voice, agrees with this conclusion, though he may well recoil from my way of arguing for it.

48. Although practical judgments are decisions, I can theoretically judge that a decision I am making or an action I am performing meets its internal standard of correctness. It is not like my present activity is in a blind spot. Whether I possess this

evaluative thought.⁴⁹

These reflections are but a start to an account of practical reason as a capacity of distinctively practical cognition wherein acting itself is an act of reason or a material process that is part of an exercise of a cognitive power. It is but a start to rejecting a wholly “contemplative model of knowledge” on which “the facts, reality, are prior and dictate what is to be said, if it is knowledge” and explaining a kind of knowledge that is “the cause of what it understands” rather than being “derived from the objects known” (Anscombe [1957] 32; 53). A full account must show that practical knowledge unifies knowledge of what is good for me to do with knowledge of what I am doing such that a robust standard of correctness governs our exercises of this capacity. It must respond to a variety of objections to the particular view of practical reason as our kind of will and the general constitutivist view of capacities in terms of which I explain practical reason. The need for this account is clear, though, only when we understand and reject intellectualism. We must reject this intellectualist legend, for it is a myth unfit to account for the capacity that it is invented to explain. Only then can we understand practical reason. Only then can we understand ourselves.

theoretical knowledge, though, is external to the exercise of practical reason and need not be in place in order for me to act well. Uttering ‘I should ϕ ’ thus might express a theoretical representation or a practical representation. It no doubt expresses both in normal cases. This is not unique to the first-person case. Uttering ‘you should ϕ ’ no doubt expresses both a theoretical representation of what *you* should do and a practical representation of what *I* am to do in any normal case, even if I am just to avoid doing stuff that will keep you from doing what you should do. Still, in both cases, the acts of thought are distinct, and my claim is about distinct kinds of thoughts, not about utterances.

49. In this way, although I share a starting point with Gibbard, at least in *Thinking How To Live*, we diverge radically. We agree that practical reason is the will and that first-person present tense practical thought has explanatory priority over other normative judgments. However, while I think that practical judgments are decisions and other normative judgments are theoretical judgments about decisions, he thinks that all normative judgments are decisions. Gibbard thus shares with the intellectualists the goal of fitting all normative thought into one mold, his of a practical representation, theirs of a theoretical representation. This aspect of his view is the site of much debate and the source of much doubt. I think that I can account for the productive nature of practical judgments while explaining the evaluative nature of other judgments about what people should do in a way that he cannot. I thereby think that he, like the intellectualists, misses the differences between practical judgments and other normative judgments in his quest to explain the way that these kinds of thought fit together.

THE EMPTY I

Since the man of common sense makes his appeal to feeling, to an oracle within his breast, he is finished and done with anyone who does not agree; he has only to explain that he has nothing more to say to anyone who does not find and feel the same in himself. In other words, he tramples underfoot the roots of humanity. ... The anti-human, the merely animal, consists in staying with the sphere of feeling. Hegel *Phenomenology of Spirit* §69

1. WHY, WHAT, AND HOW

What makes up a practical thought, an answer to the question ‘How am I to act?’ It must include what I am to do. After all, I am hardly acting if I am not doing something, even if that something is ‘doing nothing’. It also must include how I am to do it. After all, I need to exercise my skills in situations not entirely of my making and not entirely under my control in order to do anything at all. I need to know how to do what I am to do given my situation and skills. It must also include why I am to do it. After all, I do not act without a goal in mind, like a wayward rocket that goes until it runs out of juice or into something formidable. I act in order to achieve something, even if that something is just ‘doing anything’ or ‘doing nothing’.⁵⁰ In a practical thought, I represent what I am to do, how I am to do it, and why I am to do it. These three aspects are not randomly put together. Instead, the ‘what’ depends on the ‘why’ in the sense that the ‘why’ gives the end and the ‘what’ provides the means to that end. The ‘how’ depends on the ‘what’ and thereby depends on the ‘why’ in the sense that the ‘how’ specifies the means given my situation and skills. In a practical thought, then, I represent a *way* to take sufficient *means* to my *end*. I represent an action.

50. This claim does not conflict with the fact that sometimes we do something ‘for no reason’. After all, what we do ‘for no reason’ invariably itself has a means-end structure. Kicking a rock for no reason includes swinging my foot back and then forward in order to kick the rock. Jumping into a puddle for no reason includes jumping in some specific direction in order to land in the puddle. The issue about whether I can act ‘for no reason’ is thus not about whether every action that I perform has a means-end structure or, more accurately, a part-whole structure that maps onto the means-end structure of the practical thought that it realizes. What it is instead about is unclear to me.

What does reason contribute to this representation or, what is the same thing, what is practical reason? In this chapter, I shall argue that *instrumentalism* about practical reason in its simple and sophisticated forms cannot account for these aspects of practical thought. Let me here say something about instrumentalism, simple and sophisticated, before I explain my criticism of it.

Simple instrumentalism says that the only role for reason in exercises of our capacity to act is to form means-ends beliefs on the basis of ends generated through another capacity. Put in terms of the three aspects of practical thought, something else answers the question ‘Why am I to act?’, and reason then constructs beliefs about what to do and how to do it in order to achieve that end. On this view, an exercise of practical reason just is an exercise of theoretical reason about how to achieve an end that a capacity distinct from practical reason establishes.

This simple view is implausible. Practical thought is about the ‘why’ as much as it is about the ‘how’ and ‘what’. An account of practical thought that prohibits reason from contributing to this aspect of a practical thought has two problems. For one thing, it conflicts with the first-person perspective of agency, in which my end is as up to me as anything else. For another, it makes a practical thought look like a heap of representations from different sources, with questions about the relationships, both metaphysical and normative, between the capacities that provide these aspects. Each problem is formidable, and together they create quite the challenge to instrumentalism.

Simple instrumentalists can stand their ground. More promising, though, is for them to argue that we settle the why of our action in the same way that we settle the what and the how. After all, the core thought of instrumentalism is that an exercise of practical reason always starts from something given to it by another capacity. The work of practical reason is to find a way to promote, realize, or

achieve that something. Although one task is to find means to that end, another way to secure an end is to make sure that I value other stuff that harmonizes with it and do not value stuff that conflicts with it. In a sense, an end structures an evaluative standpoint against which I can measure many things, whether means, other ends, social situation and institutions, and whatever else might have normative relevance with respect to the end. In this way, I can use a general end that I get from elsewhere in order to set specific ends of particular actions, or so the sophisticated instrumentalist claims. Whether about the why, the what, or the how, then, instrumentalists think that when I answer a practical question, I “treat some [values] as fixed, if only for the time being [I do] not stand completely apart from [my] starting fund of [values]: rather, [I use] them, reason[] in terms of them” (Street [2006] 124).⁵¹ The value is the basic normative source for the end, the means, and the way. It constitutes my agential standpoint with respect to this practical thought. All such thought is in effect thought guided by the value in the way that means-end thought is thought guided by an end.

I shall argue that this view cannot explain the self-consciousness of our action because it cannot show that a practical thought has a standard of correctness. In brief, I argue that it can show that a practical thought has a standard of correctness only if it can show that there is a *correct* set of values to use in order to answer a particular practical question. After all, on this view, something counts as a good

51. Throughout this paper, I use ‘values’ for the attitude in question for the sake of rhetorical consistency. For example, in this passage, I substitute ‘values’ for ‘evaluative judgments’. Sharon Street uses a number of different terms to refer to the same thing. She says that “*normative judgments* and *judgments about reason*” are interchangeable with each other and with claims that use “the language of *value*, *should*, *ought*, *goodness*, what *makes sense*, what’s *rational*, *worthwhile*, and so on” (Street [2008] 209n.5). She also uses valuing something and “tak[ing] or judg[ing] this ... to be valuable” interchangeably (Street [2010] 366). She distinguishes “*valuing*, or *taking something to be a reason*, or *normative judgment*” from the attitude of “being pleasantly attracted” to something that she calls “mere desire” (Street [2012] 43-4). See also Street [2008] 227-30. She also distinguishes this attitude from belief (Street [2008] 230; [2010] 376). I choose to use ‘values’ throughout because Street and Simon Blackburn distinguish valuing from mere desiring in terms of certain ‘constitutive requirements of valuing’. See (Street [2008] 227-9, 230n. 38, 242n.57; Street [2012] 42-4) and (Blackburn [2010] 310-1). Since nothing in my criticism turns on the differences between desires, values, and whatever else you might invoke, this regimentation does not matter.

answer to a practical question because it accords with the values I use in order to answer that question. In order to consider this answer as non-arbitrary, then, I must think that which values I use in order to answer it play that role non-arbitrarily. I shall argue that I cannot explain to myself why I use some set of values rather than another in order to answer practical questions. From my own perspective, then, which set of values I use in order to answer practical questions is arbitrary. I then cannot have a sense of why I act. Instrumental thus cannot account for practical thought and is thereby not an account of practical reason.

2. ON STRANGER TIDES

Sophisticated instrumentalism might seem like an odd view. What does it mean to say that practical reason is a species of theoretical reason, distinguished from others when it makes specific judgments of a characteristically intrumental sort, and yet say that these judgments can be about ends as well as means? In this section, I shall answer this question with an eye towards highlighting the aspects of the view that I shall criticize. Let me start with an analogy.

Think about logical principles. Although we do not call them instrumental, these principles are like the instrumental principle because I cannot come to know things simply from them.⁵² I instead must first gain knowledge from elsewhere. These principles let me extend that knowledge. They are like functions that only produce an output when given an input from elsewhere. In other words, the capacity to infer is not a self-sufficient capacity to know because any series of inferences must proceed from an original bit of knowledge provided by another capacity to know. It is instead a capacity to know

52. I am ignoring reflective knowledge of these principles. That is knowledge of them, not know from them. I also am assuming that the perfect case of inference is coming to know something from something else that I know and that the perfect case has explanatory priority over the imperfect case. Nothing in this paper turns on this assumption.

something on the basis of something else that I know, and a being can thereby have a capacity to infer only if she has at least one other determinate capacity to know. Or, what is the same thing, a theoretical inference is an exercise of theoretical reason that extends truth on the basis of true theoretical thoughts that bottom out in an exercise of theoretical reason of some other kind.

In the same way, the instrumental principle characterizes a capacity that is not self-sufficient. It characterizes a capacity of practical inference. It is a capacity to move from an input that another capacity provides to outputs. In the paradigm case, I move from a representation of an end produced by another capacity to representations of what to do and how to do it in order to achieve that end. If practical reason is a species of theoretical reason distinguished from others because its judgments about what to do and how to do it in order to achieve an end given from elsewhere, setting ends thus seems to be outside of the purview of practical reason. How can it be otherwise? A practical inference is then an exercise of practical reason that extends goodness on the basis of other good practical thoughts.

According to the instrumentalist, though, practical inference differs from theoretical inference in that it bottoms out in an exercise of another capacity altogether, not an exercise of practical reason of some other kind. Instrumentalists posit this difference between practical reason and theoretical reason partly because for them an exercise of practical reason just is an exercise of theoretical reason about specific objects. Since they do not think that we have theoretical knowledge of the good, the ends must come from a different capacity.

Or so it seems, and so, I shall eventually argue, it is, to the detriment of instrumentalism. The main thought behind sophisticated instrumentalism, though, is that the same activity that generates representations of means from representations of ends can generate ends, appearances to the contrary

notwithstanding. The trick is to understand evaluation of ends as taking place in terms of other ends. I shall explain this idea first in terms of familiar analogies and then more abstractly.

Although instrumentalism is a view of practical reason, that fact might surprise you. Instrumentalists so often present the view in terms of nautical analogies that you might well think that it is really about ill-fit sea-faring vessels. The most common analogy is with a non-swimmer at sea without a life jacket floating on a raft with interlocking planks. Although I need not stand on any particular set of planks, I need to stand on some planks or other in order to not sink and die. So long as they fit together and have not developed sufficient rot so that they cannot support my weight, any will do. Instrumentalists claim that I likewise must 'stand on' some of my values in order to answer any practical question. I must use at least some of my values, but I need not use a particular set of my values in this way. Any non-empty subset of them will do. Without using one of them, though, I cannot answer practical questions. I then cannot act.

Simon Blackburn offers a different nautical analogy that is in two ways better and in one way worse than the raft analogy. According to him, the practical standpoint is like a ship and my values are like crewmembers who each prefer a course. Given a crewmember's preferred course, some routes count as good and others count as bad. Sometimes a crewmember takes the wheel and steers the ship for one horizon. Sometimes a different one takes control and sails a different way. Sometimes crewmembers join together because they want to sail in the same direction. They cajole others to join them, overrun and disregard them, or even make them walk the plank. Maybe sometimes pirates board the ship, get some of the original crew to join up through enticement or Stockholm Syndrome or what have you, and execute the rest. Regardless of the details of which crew-member sets the course, though, at least one of them

must have control if the ship is not to drift aimlessly in the waters.

Blackburn claims that practical reason has a similar structure. A value specifies some answers to practical questions as good because they in some sense accord with it. It specifies some answers to practical questions as bad because they in some sense discord with it. Some of my values might join together when the same answer accords with each of them. Or some values might conflict with each other, and then something must happen so that I can answer a practical question. Values might get revised or abandoned. Or maybe I have a conversion experience where I come to have a value that does not come from my previously held attitudes. Regardless of the way things actually go, though, instrumentalists say that I must answer practical questions according to the criterion that some non-empty subset of my values constitutes. Values are the central explanatory element with respect to exercises of practical reason in the same way that crewmembers are the central explanatory element with respect to the course of the ship.

This analogy is better than the raft analogy because it highlights the way that practical reason is part of human action or is primarily a matter of answering the question 'How am I to act?' and acting on the basis of that answer. The raft analogy makes practical reason seem as if it is fundamentally a matter of evaluating my values from the standpoint of my other values. However, human life is not quite that self-centered, at least not in that way. I evaluate my values because I use them in order to figure out how I am to act, and I will only act well if I act from the correct values. The raft analogy obscures this point since it is not as if which planks I keep in the raft determine its route. The planks only determine whether I sink, and any planks that fit together well enough will sustain me. Human life is about more than this concern for mere psychic coherence explains.

The ship analogy also highlights how instrumentalists think that values play a *constitutive* role in determining whether something is a good or bad answer to a practical question. In the raft analogy, although I must stand on at least one of the planks in order to check out other planks, the set of planks on which I stand do not determine whether the other planks should stay or go. They do not determine what counts as rottenness in the other planks, at least on any raft with which I am familiar. The ship analogy is better than the raft analogy because the goals of the crewmembers determine whether a particular route is a good or bad answer to the question ‘In which direction are we to sail?’

This ship analogy is one sense worse than the raft analogy, though, because it does not show that there is any agent involved. In the raft analogy, the agent is identical to the person on the raft. She uses the planks but is not the planks in the way that an agent uses her values but is not her values. Where is the agent in the ship analogy? Although he claims that the “person is the totality composed of body and form, ship and crew”, Blackburn never mentions the ship when he spells out the analogy (Blackburn [1998] 251). I take it that the point of it is to capture how my attitudes can interact with each other ‘immediately’, but they can only interact with your attitudes ‘mediately’ through speech or writing or physical intervention or whatever. That is why mine are mine and yours are yours, or why I am one person and you are another. Similarly, the crewmembers of one ship can engage with each other ‘immediately’. They can just shout at each other, shoot each other, punch each other, and all that jazz. Crews of different ships can instead only engage with ‘mediately’. Their ships must get close enough for the winds to carry their voices and bullets or for members of one ship to board the other. That part of the analogy might distinguish you from me, but it only does so if the analogy has an account of agency that captures the me and the you that must be distinguished. The problem with this analogy, though, is that

there is nothing for the agent to do. The agent is not *one of the crewmembers*, since each crewmember has equal claim to stand in for the agent, which is at best an analogy for madness, not human life. Whence the agent?⁵³

Of course, analogies are analogies, not tautologies, and perhaps all illustrative ones lack some relevant details. More important than what they leave out, though, is what they build into the picture of practical reason without mentioning it. Each analogy, in its own way, seems to show how to stay true to the spirit of instrumentalism while accounting for our ability to set ends. I can ask and answer the question about the end of action by using some of my values in the way that I can ask and answer the question about the means and the ways of actions. Just as I evaluate possible means and ways to do things from the 'standpoint' of the end, I evaluate possible ends from the 'standpoint' of my values. In each case, practical reason simply evaluates something in light of something else given to it through the exercise of another capacity. In fact, in each case, the standpoint of evaluation and decision is really the standpoint of my values since it is not like I opt for means that conflict with my values when they happen to serve my end.

Why is such evaluation instrumental, or at least recognizably tied to the instrumental principle so that it extends instrumentalism rather than controverting it? To start, in each case, the judgments in question are just judgments about the relationship between a set of values taken as given and other things, whether means, ways, ends, or other values. Moreover, the type of evaluation in question is the

53. One way to respond to this issue is to treat the claims about the way values and beliefs function in order to produce action as part of a metaphysical reduction of agency and action. On this view, to exercise practical reason just is for my values to produce bodily movements, influence my other values, and exert certain efficient causal effects on myself and the world in the same way that to be water just is to be H₂O. This kind of view is prominent in action theory. I argue against it in 'I do. What Happens?' Instrumentalists, though, do not address any of the questions that a metaphysical reduction of agency must answer. Hence, I assume that they intend to offer an account of self-conscious action at the agent-level, which is the only sense I can make of an agent 'using' values and 'treating' some of them 'as fixed'.

same in each case. After all, think about the crew of the HMS Blackburn and what it means for the nature of a value. Each crewmember has their own destination in mind. They evaluate courses and other crewmembers only in light of whether the courses and other crewmembers might help them reach that destination. Most importantly, though, each crewmember is after the wheel, since the only way to reach the destination is to point the ship in the right direction. Putting these points to work in an account of the nature of a value, values differ from each other on the basis of what is valued. Each value is 'after' what it values, whether by trying to promote it, realize it, protect it, and anything else that might count as going for it. For a value to 'take the wheel' is for it to be in control of our practical thought and action such that we privilege going for it over other things. Part of taking the wheel is keeping other values from getting in the way, particularly other conflicting values. Hence, one part of going for what is valued is for this value to serve as the evaluative criterion by which I not only determine what to do and how to do it but also by which I determine why to do it and what other things to value. That is to say that controlling practical thought is a partial means for any value to realize its object or achieve what is valued. In all of these roles, the value is given, and reason works out what is likely to lead to achieving what is valued. That is why sophisticated instrumentalism is a version of instrumentalism.

3. *EXTENDING GOODNESS*

Why be an instrumentalist? At least one thought behind the view is about the place of reason in our lives. Reason does not generate its own content, according to instrumentalists. I do not know of a non-question-begging way to present this view, but it does offer a superficially adequate account of practical thought. Let me run through it before I criticize the instrumentalists in the next section.

Both nautical analogies seem to show that values matter because without them I would have no basis to answer practical questions. Indeed, Street and Blackburn invoke this idea of emptiness in their presentations of their views. To see this point, think about the fact that everything in the world falls into two mutually exclusive and jointly exhaustive classes. Some things I value, which include things that I 'disvalue' in the sense that I value their destruction or value avoiding them and the like. I fail to value everything else. They do not matter to me because I am ignorant of them or because I do not care about them. I neither value nor disvalue them. I am indifferent.

In order to answer a practical question, I need some way to discriminate good answers from bad ones. What provides that criterion? A principle of practical reason like the instrumental principle will not help. It can only provide answers to questions if it gets content from elsewhere—in the paradigmatic case, if an end is provided, and in the sophisticated cases, if a criterion on the basis of which to evaluate things is provided. I surely will not use the things that I do not value at all. If I am ignorant of something, I cannot base an answer off of it. I also cannot base an answer off of something if I am aware of it but indifferent to it. How would I understand myself in this activity, choosing on the basis of the indifferent? If I am indifferent to something, whether things promote or frustrate it does not matter to me, and goodness and badness do not get a grip here. Hence, it seems that the only things that I can use as the basis for my answer to practical questions are my values. They pick out what matters to me, and something counts as a good answer to a practical question to the extent that it promotes, realizes, or facilitates that value. What is the alternative? The analogies are designed to make it seem like the answer is nothing. After all, the only alternative to standing somewhere on my raft is drowning, and the only alternative to one of the crewmembers taking the wheel is the ship drifting aimlessly.

You might wonder whether principles of practical reason in some way determine whether something is a good answer to a practical question. A Kantian view of practical reason, for example, claims that the constitutive principle of practical reason is able to provide its own premises in a way that the deductive principles of theoretical reason cannot provide their own premises. On this view, practical reason is a self-sufficient capacity in a way that the capacity to deductively reason is not. Instrumentalists, though, do not think that this kind of view is possible, not so much because they understand it on its own terms and reject it but because they interpret it in terms of their view of our agency and then declare it nonsense. Because it helps clarify instrumentalism, let me explain this idea with respect to both Street and Blackburn.

Street first. Technically, as far as I can tell, Street does not think that there are *any* principles of practical reason, let alone any principles of practical reason that provide their own premises. This claim might seem odd since just about everyone thinks that the instrumental principle or something like it is a formal principle of practical reason. Street, though, understands the phenomena that lead other philosophers to posit something like the instrumental principle in a different way. According to her,

it is constitutive of [valuing] *Y* that one also, when attending to the matter in full awareness, take oneself to [value taking] what one recognizes to be the necessary means to *Y*. One *cannot* take oneself to [value] *Y* without taking oneself to [value] the means to *Y*, where the force of the *cannot* here is not rational ... but rather analytic or conceptual ... [This is a] purely formal statement[] about what is involved in the very attitude of [valuing]. [It] make[s] no substantive assumptions about what [values] there are; [it] merely state[s] what is involved in [valuing something] in the first place. If someone 'violates' [it], then she is not making an *error*; she is merely not [valuing] anything. (Street [2008] 228-9)

As this passage fairly clearly states, there is no instrumental principle of practical reason in the sense of a principle that I *follow* whereby I move from one value to another. I do not value something, form a belief

that acting in some way is a means to realizing a value, and then value that means on the basis of my recognition of the relationship between acting in that way and the thing I value. Of course, the claim that I necessarily value the means to things that I value ‘non-instrumentally’ is a little delicate to state. The idea, I take it, is that it is part of the attitude of valuing something that I value taking the means to achieving or promoting that value, where ‘the means’ does not pick out *the particular things that are means* but rather means something like *whatever happens to be the means*. When I form a belief about particular means, I give determinate content to that second value. I now value a particular substantive version of the general act description that I have always valued simply in virtue of the fact that I value something else non-instrumentally.

To the extent that there are any principles of practical reason on Street’s view, then, these principles do not generate any content on their own. They simply say that to value something non-instrumentally brings with it a generic valuing of the means to that thing, whatever they might be, where this claim is a metaphysical claim about the nature of an attitude or mental state properly characterized in modal terms. Principles of practical reason cannot generate their own premises, then, because these principles only get going once an agent already values something. Hence, Street says that logical

and instrumental ‘requirements’, as these govern practical reasoning, are explained not as substantive values, but rather as features constitutive of the attitudes of valuing. To ignore these ‘requirements’ in full consciousness of what one is doing is not to make a *mistake* about a normative matter; it is merely to fail to value. Apart from the attitude of valuing, there *is* no such thing as a mistake. Either one values things or one does not ... (Street [2012] 40)

Because of this understanding of the principles of practical reason, Street does not take the Kantian view of practical reason to be the claim that the principles of practical reason generate their own content. She

instead understands it as the claim that all values in some way imply a specific value.⁵⁴ On this view, Kantians about practical reason are instrumentalists who just happen to think that one particular value has a special pride of place because of its inferential connections with every other possible value. In fact, given Street's understanding of the nature of principles of practical reason in light of her instrumentalism, it is not clear that Kantianism could be anything else.

Blackburn likewise seems to treat instrumental requirements as non-normative conceptual or metaphysical claims about the nature of the mental states that interest him. For example, he says that the claim that normative properties, or at least the property of goodness, supervene on natural properties "becomes the requirement that you endorse things in the light of their other properties: one of the very few, and perhaps the only, constitutive requirement of evaluation" (Blackburn [2010] 310). This language of 'requirements' sounds like a normative claim about the way to evaluate. However, since he elsewhere says that "a constitutive rule ... tells us *what it is*" to have the particular mental state it concerns, the claim about the constitutive requirement of evaluation in fact sounds like a metaphysical claim about the nature of whatever mental states explain evaluation. (At least, such a reading is required unless he endorses something similar to my constitutivism. I somehow doubt that he would.) The instrumental principle might seem to warrant a treatment as a conceptual truth about the nature of values similar to Street's view. Yet there are two problems with this thought. For one thing, the claim that supervenience is 'perhaps the only' requirement on evaluation seems to militate against that

54. With justification, since Street's discussion of Kantianism is really about Christine Korsgaard's argument in *The Sources of Normativity* that valuing anything whatsoever commits me to valuing my humanity and, in turn, everyone else's. This view seems to me to represent what Kant's view looks like if you try to understand it within a broadly Humean or empiricist moral psychology, not his view on its own terms. I also doubt that it is in the end the way to understand Korsgaard, at least if this view is to be continuous with the view in *Self-Constitution*. These issues are off-stage here, though, so I leave them.

reading. It is not like the instrumental principle is an obscure topic that might escape his notice. For another, he claims that “failing to adopt means to ends may be just a defect of passion” (Blackburn [1998] 239). The ‘may’ here sounds like it qualifies the normative significance of this failure, which jibes with his claim to not “know whether it is useful to talk about defects of rationality” even when someone has only one end, understands some means as essential to that end, and yet does not take them (ibid. 320). Yet can defect coexist with a constitutive rule that tells us what something is? Without something like my constitutivist background, it is hard to say yes. Yet Blackburn offers nothing like a constitutivist story of the basis of normativity anywhere with which I am familiar, and the project seems antithetical to his general philosophical ambitions.

Whatever you make of that stuff, Blackburn like Street understands the Kantian view of practical reason as a claim about a particular value, and it likewise follows from the way that he presupposes instrumentalism in his presentation of the view. Blackburn presents Kantianism in terms of his nautical analogy, where the Kantian claims that there is a captain among all the crewmembers who has authority over the course of the ship. Presumably, this captain must be different from all of the other persons on the ship in a way that explains this authority. What can this difference be? In the analogy, the crewmembers stand for my values. They represent anything that I value, whether positive or negative. The captain cannot stand for something that I value, then, because he would then just be another member of the crew, not a captain. But what else might it be? Something that I do not care about at all? That hardly can have the authority to determine my answer to a practical question. Recognizing that a Kantian means to reject the idea that any *substantive value* can play the role necessary in order to explain the possibility of human agency, Blackburn thinks that they can only think that an *insubstantial value*

plays that role. But an insubstantial value is no value at all, and so Kantianism is no view at all. If a Kantian captain is in charge, according to him, then it “is indeed not clear whether these particularly blissful ships travel anywhere” (Blackburn [1998] 246). This claim is of course nonsense. A ship with no guidance does not sit still. It drifts aimlessly, traveling wherever the wind and current take it until it washes up ashore. The point, though, is clear enough.

Although neither of these arguments explicitly presents a rationale for instrumentalism, the thought behind them is in my opinion clear and part of a familiar general view of our agency and mental life. The work of reason, on this view, is not to generate content. It is instead to work on content supplied to it from other capacities in the same way as a capacity to infer works on content supplied to it from elsewhere. With respect to a capacity to infer, the content comes from perception and other capacities to know that let us take in things from the world. With respect to practical reason and the instrumental principle, nothing in the world can provide the relevant content. For one thing, these theorists have familiar epistemological and metaphysical qualms about this kind of receptivity to these normative facts. For another, and more importantly, the content on which reason operates in the practical sphere must play a role not only in guiding thought but also in guiding action. A mere extension of truth across inferences might seem ill-suited to play that role, and a receptive faculty to normative facts might seem like it only provides content for that type of extension of truth. Values instead seem like the kind of thing that can simultaneously provide the content for exercises of reason according to the instrumental principle and also the guidance of action. In the former role, a value divides up aspects of the world, including ourselves, into the good, the bad, and the neutral depending on whether it tends to help promote or realize the value in question. In the latter role, it provides guidance of action since to

value something is to be set to promote or realize it when possible, other things being equal.

Try it this way. According to many people, anyway, one distinction between theoretical reason and practical reason is that the world has a kind of authority over us in exercises of theoretical reason in a way that it does not in exercises of practical reason. Theoretical judgments answer to the world in a way that practical judgments do not. This contrast is coarse, with qualifications needed in both directions. Exercises of theoretical reason are not purely passive, even in paradigm cases of passivity like perceptual knowledge. Exercises of practical reason answer to the world in all sorts of ways since we act in a world over which we lack complete control. Suitably qualified, though, the basic thought might seem to express how practical reason involves a kind of self-determination that theoretical reason does not. One way to get into the idea behind sophisticated instrumentalism is to think of values as accounting for this self-determination. After all, if they constitute the criterion by which I evaluate, deliberate, and decide, I need not get content from elsewhere in exercises of practical reason in the way that I do in exercises of theoretical reason. In some sense, I provide the content, not in the sense that the content is about me but in the sense that my values determine what is relevant and what is not. Hence, positing this role for values might seem to explain one of the salient differences between theoretical and practical reason.

Again, I do not mean here to present a non-question-begging argument for instrumentalism. I instead mean to show that the view offers an account of our life, practical and theoretical, where the parts seem to fit together. I shall argue against this account, on grounds largely internal to instrumentalism. I can, though, see the way into its view of ourselves and our place in the world.

4. CONFLICT AND CHAOS

In the rest of this paper, I will argue that if sophisticated instrumentalism is true, I must regard my answers to practical questions as arbitrary from a normative point of view. This arbitrariness shows that values cannot explain this unique kind of self-determination and thus undermines the possibility of instrumentalism accounting for practical thought.

In order to get the problem into view, think about how instrumentalism must understand conflicts between values. As a finite agent, some of my values will be inconsistent with others in obvious and in subtle ways. Some obvious conflicts border on the incoherent, such as valuing and devaluing the same thing in the same respect at the same time. Other obvious conflicts are just a part of our life, such as when I value two things that in principle can be mutually realized but in practice cannot because of contingent facts that are not up to me. Institutional structures, the lack of time and energy, and all the other marks of human life might make it impossible for me to realize two values. While these cases are interesting, most important for my purposes is a kind of subtle conflict built into instrumentalism such that every value conflicts with every other value. Let me explain.

Part of what makes sophisticated instrumentalism *instrumentalism* is that it treats evaluation of values and possible ends of action on the same model as evaluation of means in light of an end. I evaluate from some set of values taken as given, not from something outside of the totality of my values, even if I am evaluating some of my values. Every value, though, in a sense aims at the promotion, protection, and realization of its object. Hence, just as every value represents taking the *means* to realizing its object as good, every value also represents its playing the role of constituting the standard by which I evaluate as good. After all, I shall not take the means to realizing the object of *this* value if I do not set an appropriate end, and I will not set an appropriate end if I use some other value as the basis for evaluation. That is to

say that playing this role is a partial means of the realization of the object of any value whatsoever, at least so long as instrumentalism treats evaluation on the model of following the instrumental principle. If one value plays this role, though, others cannot. In this sense, every value conflicts with every other value. To return to the metaphors, every value says for me to stand on it. Every value says that it is to take the wheel.

How do I resolving a conflict in values? This question is a practical question, which means that I must answer it from some standpoint that some of my values constitute. Which one? This question itself is a practical question of the form ‘How am I to answer that first practical question?’ or, more literally, ‘Which subset of my values am I to use to answer that first practical question?’ I then need to use a subset of my values in order to answer this question. How am I to figure out which subset to use in that instance? You see where this is going. I shall here argue that the instrumentalist cannot answer this question.

Some natural answers to other kinds of conflict obviously cannot solve this case. For example, in a more ordinary case of conflict between two values, one of these values might conflict with many other values. Perhaps it is even the case that this value conflicts with all of my other values. The natural thing to say in this case is that I should not use this value in order to answer the first practical question. I should instead revise this value in order to make it consistent with my others. If consistency is impossible, I should abandon it. The sophisticated instrumentalist cannot use this strategy in the current case, though, for two reasons. For one thing, each value says of itself that it should be in charge of evaluation. In this way, every value conflicts with every other value, and we do not get the lopsided structure of the other case. More importantly, even if a number of my values are in some sense ‘on the same side’ of this issue

as against another value, that fact is irrelevant. Why should it decide the correct answer to my practical question? According to instrumentalism, a fact of this kind is only relevant to this practical question if the criterion that structures my evaluation says so. Whether this criterion says so depends on which values establish it. That is the question that I am trying to answer in the case of a conflict of values. The value that is not on the larger side will, no doubt, count this fact as counting against letting the larger side rule. After all, its role is to realize its object, which requires it governing evaluation. According to it, whether a fact is relevant to a decision and in which way depends on whether it helps or hinders realizing the valued object. In this case, trumping for the majority gets in the way of realizing that object. Hence, the fact of the majority counts against the majority ruling from the standpoint that the value in the minority constitutes. It counts for the majority ruling from the standpoint that the values in the majority constitutes. This is just another conflict.

I shall push this point repeatedly in what follows, so let me run through it one more time, somewhat more abstractly. Consider Blackburn's claim that

we can usefully compare the ethical agent to a device whose function is to take certain inputs and deliver certain outputs. The *input* to the system is a representation, for instance of an action, or a situation, or a character, as being of a certain type, as having certain properties. The *output* ... is a certain attitude, or a pressure on attitudes, or a favouring of policies, choices, and actions. Such a device is a function from input to output: an ethical sensibility. (Blackburn [2008] 5)

As far as I understand this view, the way things go is that a representation of a fact is an input that is in the first instance devoid of normative relevance. Whether and in what way it is relevant depends on its relationship to my values that constitutes the function or device that produces an output given the input. Facts are relevant to the the question 'How am I to act?' only given the values that constitute my practical standpoint at a time. I determine my end of that basis. I then can extend 'goodness' across

inferences, determine the means and the way to pull them off.

Conflicting sets of values, though, differently determine the normative relevance of any fact that distinguishes between them. Each value says that the means to realizing its object are good and things that get in the way of realizing its object are bad. Each value thereby says that anything that sides with it governing evaluation is good and anything that sides against it is bad. There is thus no way to resolve a conflict between values. Anything that might distinguish one value from another will count in one way by one of their lights and in the other by the lights of the other. To use the language of Blackburn's input-output device analogy, it is part of the input side of the equation, not part of the device. Different parts of that device count this input differently, and nothing privileges one part of the device over another.

The structure of instrumentalism is the problem here, not anything about the details of the case. To see why, consider the following response. What if I in fact have values that concern how to resolve cases of conflict? Street claims that "even though we don't always articulate them, we all implicitly endorse innumerable such [values] about the proper trade-offs between different sorts of values" (Street [2008] 233). What if one of my values about trade-offs is relevant to this case because it explicitly concerns the way to 'trade-off' involving these values or because it is a general principle like 'always side on the side of the action that coheres with most of my values'?

Yet this response just adds one more value to one side of the original conflict. The original problem does not stem from the number of values or the particular content of the values. It stems from the fact that some of my values are in conflict, regardless of which values or the number of them. Each of them says that it should govern evaluation. In a way, then, the idea of values about trade-offs is nothing

new and nothing added. Every value is itself partly about tradeoffs since each value says that it should govern evaluation and thus that I should resolve every instance of conflict in its favor. Adding values that solely concern trade-offs thus just adds more values to the conflict, with the same standing as any other value. After all, they determine the correct answer only if they govern practical thought. Why should they govern over the others?

Street recognizes this kind of problem, or at least she recognizes an issue that leads her to a view that also might appear to respond to this kind of problem. She claims that when

we're asking what reasons a given agent has all things considered—and not just what reasons she has from the standpoint of some (implicitly or explicitly) specified subset of her values—which standpoint gets priority? The answer, roughly, is that the standpoint that determines what reasons she has is whichever standpoint is most deeply *hers*, where this is a function of how strongly she holds the [values] in questions and how close to the center of her total web of [values] they lie. (Street [2008] 234-5)

She explains that

the priority according these [values that lie close to the core of a person's interlocking web of normative judgments] doesn't reflect a substantive value, but rather reflects the fact that we are asking about *agent A's* reasons, not someone else's reasons, and agent *A* is, in an important sense, to be identified with her most strongly and centrally held values. (Street [2008] 235n.45)

According to these views, there is a metaphysical account of personal identity at a time that is, essentially, a practical version of Quine's web of beliefs—the web of values. This web of values explains the way to resolve conflicts between my values. I just need to calculate the strength and centrality of the values on each side of the conflict and figure out which side has more total strength and centrality.

Assume that the ideas of a web of values and centrality make sense. The problem with this response is the same as the problem with the earlier responses. The strength and centrality of a particular value relate to others are normative properties of it or they are not. Start with the idea that they are not,

as Street seems to claim. They then have no intrinsic normative significance. They are on the input side of the practical agency function. They are normatively relevant if and only if some value that partially constitutes the practical standpoint says that they matter to the practical question at issue. These facts then cannot explain the way to resolve a conflict. In a case of conflict, different values give different verdicts to whether and in what way these facts are relevant depending on whether they partially lead to realizing the object of that particular value. Just as the free spirit sees the overgrowth as a good aspect of taking the road not taken but the square sees it as a bad aspect, so the strong central values see their strength and centrality as supporting their claim to govern and the weak outsider values see it as undermining that claim. Nothing normatively relevant distinguishes these values from each other because each value is competing to determine normative relevance. Strength and centrality then are unfit to solve the problem of conflict if they are non-normative properties.

Yet the same problem arises if the claim that the strong central values should rule is a normative claim. After all, if it is, the source of its normativity must be some value. Whether it is one of the values already involved in the conflict or a value that only concerns the centrality of values in the web is neither here nor there. It is just another value in the conflict, on a par with others in that respect and with respect to its authority. Why does this value have authority to govern practical thought? What does it have that the others lack? All of those values lump to resolve the conflict in one way or the other as well. What makes this resolution differ from any other? What answers that practical question? In a certain sense, nothing. In another sense, everything. Therein lies the problem.

Before I move to the next section, let me point out that this problem highlights one reason why instrumentalists like Street and Blackburn opt for a view of the instrumental principle wherein it is

descriptive, not normative. If the principle is normative, then its normativity must derive from some value or other. Yet there does not seem to be a value that could explain why this principle is normative in general. Each particular value demands that I take means to achieving its object, but nothing says that in general I am to take means to ends. Moreover, if something in general did, it would simply be another value among others, with no difference in standing. The instrumental principle, though, is not one substantive value among others. Its generality—the way that it always applies regardless of the value in question—rules out that view. Instrumentalists thus need a different account of it. Yet they cannot have a different account of it and say that it is normative while also holding that values are the source of all normativity. Hence, they must deny that it is normative.

5. EVALUATION AND ARBITRARINESS

Although the problem of conflicts might at first seem to arise only for specific subsets of values, it in fact arises anytime an agent has more than one value. Street thinks that “talk of truth and falsity with respect to” one of my values only “starts to get a foothold” when I have at least two values because only then is there a “standard in place to determine its correctness—in particular ... the standard set by [the] *other* value” (Street [2008] 223). Presuming that I can only ask whether to use a value in evaluation in some way if I am in position to ask whether it is correct to do so, with the possibility of correctness with respect to values comes the actuality of conflicts. What is vivid in the case of overt conflicts is in fact always present. I shall argue that this issue shows that instrumentalism is incompatible with the possibility of a standard of correctness for practical thought. It is thereby incompatible with us taking ourselves seriously in our exercises of our capacity to act. Let me explain.

Although she does not recognize the extent of the issue, Street allows for certain irresolvable practical questions. She claims that “the notion of radical choice [has] an important” role to play when “there is not a single [value] available in one’s set of [values] to settle whether some other [value] is correct or not” (Street [2008] 237). A radical choice cannot be correct or incorrect—that is, after all, what makes it radical rather than simply choice. Still, she thinks that it sets up a subsequent standard of correctness for further practical thought. In effect, in trumping for one thing rather than another, I establish it as the criterion for subsequent practical thought, at least if all goes well. Can a radical choice play this role in general, grounding the authority of whichever subset of values I use in order to answer any practical question?

No. For one thing, according to instrumentalism, a radical choice, by definition, lacks intrinsic normative significance. After all, normative relevance derives from values, and radical choice has a role to play only when values cannot answer a practical question because of conflict. However, a radical choice then cannot establish a normative standard for subsequent practical thought. Just try it. Say I have two values that conflict over which one should govern my practical thought. I make a radical choice in favor of the one. That value will say to follow the radical choice in subsequent thought, in effect saying of itself that it should govern practical thought. The other value will say to go against the radical choice in subsequent thought, in effect saying of itself that it should govern practical thought. Invoking other values, like the value of always following my radical choices, does not change anything. It just adds more values to the conflict. Adding a radical choice to the situation amounts to the same thing as adding any other fact that lacks intrinsic normative significance. It does not resolve the conflict. It just provides another occasion for it.

For another thing, and more importantly, the appeal to a radical choice shows the deepest problem with instrumentalism. A radical choice cannot be correct or incorrect. It is beyond normative evaluation, not in the sense that it is the *kind* of thing that can be so evaluated but for some reason in this case is not but in the sense that it is not the kind of thing that can be subject to normative appraisal at all. It is arbitrary. Moreover, if I make a radical choice, I know that it is in this sense arbitrary. After all, I make a radical choice because I cannot settle the issue on normative grounds. Understanding myself as making a radical choice thereby undermines my own sense of the normative standing of my practical thought that follows from it.

In order to bring out the force of this issue, let me distinguish it from something that might seem similar. Any account of practical thought needs a place for arbitrary determination because any well-formed practical thought includes one among many equally good sets of sufficient means. Take even a simple case of crossing the street. I can start off with either foot and start off at any of a number of moments. I generally do not think consciously about these things since they do not matter, but all the same I set off with one foot at one moment. In doing so, I exercise practical reason. Thus, no view of our agency can do entirely without arbitrariness in our self-determination. Unlike instrumentalism, though, this undeniable feature of our agency has a place of arbitrariness *within* a generic practical representation that itself is subject to a normative standard. It is a choice between two equally good ways of performing an action that I already represent as a good thing to do. Similarly, on most views of practical reason, an arbitrary choice between two alternative actions is a choice between two actions subject to a normative standard. It is a choice between two good actions. In contrast, instrumentalism must posit an arbitrary choice about everything. It is not just about the way to pull off the means to an end, or the means to the

end, or the end. It is also about the values that determine whether anything is good or bad.

Try it this way. I ask myself whether one value or another should govern. I cannot find any normative basis to answer this question or, what is the same thing, I find parallel normative bases running in both directions. Moreover, I know that this deadlock is not even in principle the result of an epistemic limit on my part. Nor is it an isolated case. According to instrumentalism, every answer to a practical question presupposes some set of values that constitutes the criterion that determines whether an answer is good or bad. In every practical question, multiple values with competing verdicts are relevant, if only because these values will speak to practical questions about which practical questions to even consider. I always have the question of which value to use, and nothing can settle it but an arbitrary choice. This arbitrary choice might set me off along a specific path, but I know that this path lacks any kind of normative standing. After all, I am on it because of my arbitrary choice, which I recognize as not subject to a standard of correctness and which I thereby recognize could have gone the other way. Every value will say of itself that it should govern practical thought and hence will count following my arbitrary choice as good or bad depending on whether the arbitrary choice is in its favor. To follow the arbitrary choice is, of course, to answer a practical question, done according to a subset of values. Which subset? The same issue arises, and the only way to answer it is with an arbitrary choice. At every point, then, I determine whether to go in one way rather than another by making an arbitrary choice. I thereby cannot regard any of my practical thought and action as subject to a standard of correctness.

There is then no role for values to play because whether they are to play a role is a normative question. According to the instrumentalist, I must use values in order to answer this question, but which values to use and in what way is itself a normative question. The only way to settle any of these questions

is with a descending series of arbitrary choices, about what to use, in what way, and whether to stick with it. The only kind of practical thought is then a kind of arbitrary choice. It is not subject to normative standards, and I thereby cannot regard any of my practical thought and action as subject to such standards. The capacity for arbitrary choice is in the end the sole determining force of our agency. To revert to Blackburn's analogy, instrumentalism posits the empty captain who randomly determines the heading for the ship.

If at bottom every answer to every practical question is a radical choice, though, exercises of practical reason are not subject to a standard of correctness. Think about instrumental thought. Importantly, this thought is not merely thought about causal relations, nor is it merely thought about various things that I can do. Instrumental thought is about how to achieve what I have set as my end. Thought about how to pull off an action in my circumstances is likewise about about how to perform what is in fact my means in order to achieve what is in fact my end. Moving in the other direction, according to instrumentalism, thought about ends is about why I should act given what are in fact the values governing my practical thought. In the end, according to instrumentalism, the order of explanation runs from the question about what values are to govern evaluation to the question about ends to the question about means to the question about ways. However, if the only way to answer a practical question is through a radical choice, none of those projects are possible in the relevant sense. If I determine which value governs by a radical choice, its governing lacks normative standing. If I formulate an end on that basis, this act is from something that I recognize as lacking normative standing to something that thereby lacks normative standing. The end can no more be correct or incorrect than the value that governs because whether *that value governs* and hence what constitutes the only criterion

against which the end could be evaluated is arbitrary. If the end cannot be correct or incorrect, neither can the means nor the ways, and for the same reason. To place radical choice at the foundation of practical thought is thus to undermine the possibility of a standard of correctness for practical thought.

Sophisticated instrumentalism seems to explain the unique self-determination that differentiates exercises of practical reason from exercises of theoretical reason. Whereas theoretical reason must get its content from elsewhere, practical reason can generate its own content because my values constitute the criterion by which answer practical questions. They determine what is relevant to answering practical questions. In fact, they determine which practical questions are worth asking. So it seems. Yet placing a radical choice at the basis of every act of practical reason undermines this role for values. In the end, they do not determine anything since they play that role only given a different, non-rational, non-normative act. The promise of instrumentalism is a false hope.

I DO. WHAT HAPPENS?

That seemed to me much like saying that Socrates' actions are due to his mind, and then in trying to tell the causes of everything I do, to say that the reason that I am sitting here is because my body consists of bones and sinews, because the bones are hard and are separated by joints, that the sinews are such as to contract and relax, that they surround the bones along with flesh and skin which hold them together, then as the bones are hanging in their sockets, the relaxation and the contraction of the sinews enable me to bend my limbs, and that is the cause of my sitting here with my limbs bent. ... To call those things causes is too absurd! (Plato *Phaedo* 98c-99a)

1. ANIMAL ACTIVITY

HUMAN BEINGS are animals! Like all living being, we are not merely passive in the face of the world.

Unlike a rock, some of the happenings in my life are my doings. I am active in different ways throughout my life because I have different capacities. I am active when I digest just like my cat is active when he does and my plants are active when they photosynthesize. These things do not happen to us in the way that being swallowed by the sun or being blown to bits by the bomb happens to us. Unlike the plants, though, my cat and I can act. He can leap from my bed. He can be (gently) tossed from my bed. Even if he flies through the air at the same speed along the same trajectory in each case, he is active in the first and something happens to him in the second. Just so, I can roll down the hill in order to win the race with my sweetie. I can roll down the hill as a result of an earthquake. Even if I tumble at the same speed along the same path in each case, I am active in the first and something happens to me in the second.

Action differs from digestion and other things because it involves a point of view. Human action differs from the action of the other animals because of the self-consciousness of our point of view. My action is self-conscious in the sense that I can act because of my sense of why I act. I can ask myself the question 'How am I to act?' and act from my answer to it. I can in this way realize a practical thought. Digestion is not similarly self-conscious. While I can ask the question 'How am I to digest?', I cannot

digest from my answer to it. The idea of a digestive thought does not make sense because I cannot realize it. The other animals cannot even ask these kinds of questions about digestion or action, let alone act or digest from their answers to them. In this way, we have a self-conscious will. As far as we know, only we have a will of this kind.

Because of the uniqueness of our will, philosophers of action tend to focus on the question 'What is human action?' as opposed to the more generic question 'What is action?' They approach this question with twin goals that are difficult to jointly accomplish. On the one hand, they must explain how our action is unique amongst everything that happens in the world, including everything else that the plants, the other animals, and us human beings do. Their goal is not to eliminate the self-consciousness of human action. It is to explain it. On the other hand, they must show that we act in the same world that everything else that happens. I am writing this paper while digesting my lunch as my cat paws my calf while digesting his food as my plant soaks in the sun on my windowsill. An account of our action must not make it seem out of this world. The basic challenge of action theory is to account for our action within a general metaphysics of reality.

These two goals might not seem difficult to achieve. Everything is what it is and not another thing, after all, and yet everything exists and happens in the same world. Most action theorists, though, orient their work around additional assumptions that make our action seem like an outlier. They treat action theory as an area in applied metaphysics. When philosophers do applied metaphysics, they presuppose a metaphysics that they claim exhausts the fundamental kinds of things that exist and the fundamental kinds of relations between them. They then try to show that this metaphysics can account for other putative parts of reality that do not seem to fit in it. For example, say that a philosopher claims

that everything is fundamentally water. She must show that this metaphysics can account for everything else, say, earth, fire, wind, and heart. Otherwise, she must deny that everything is fundamentally water or deny that earth, fire, wind, and heart exist. Perhaps a philosopher instead claims that everything is fundamentally atoms. She must show that this metaphysics can account for, say, molecules and solid objects. Otherwise, she must deny that everything is fundamentally atoms or deny that molecules, solid objects, and other non-atomic things exist. *That* is applied metaphysics.

Most action theorists assume a fundamental metaphysics that seems like it is incompatible with human action.⁵⁵ They thereby address themselves to a more specific version of the basic challenge of action theory. The issue for these *causal theorists* is that there is an “obstacle to reconciling our conception of [our] agency with the possible realities [because] our scientific conception of the world regards all events and states of affairs as caused, and hence explained, by other events and states” (Velleman [1992] 129).⁵⁶ The obstacle stems from the fact that I “seem to intervene in [the flow of events], by producing some events and preventing others” when I act (Velleman [2000b] 5). In human action, “the *agent* is the source of, determines, directs, governs the action, and is not merely the locus of a series of happenings, of causal pushes and pulls” (Bratman [2001] 91). In my terminology, my action is self-conscious in the sense that I can act from my sense of why I act. The problem is that this kind of self-

55. In my view, the problem is much wider than most action theorists realize. The presupposed metaphysics is in fact incompatible with all life processes because it rules out the form of explanation essential to understanding these things. I say something at the end of this paper about a different way to proceed in action theory that draws on this thought.

56. ‘Causal theorist’ is not the best name for this group. My real target are *reductive* causal theories. Causal theories need not be reductive either because they invoke an unreduced notion of agent-causation or because they willingly invoke agential concepts in their account of the mental states that stand in an event-causal relationship to bodily movements that are my action. Although I disagree with these non-reductive causal theories on other grounds, neither matters for my purposes in this project because neither is a threat to the non-reductive account of human action that I need for my metaphysics of capacities. For criticisms of causal theories simpliciter with which I am sympathetic, see (Ford [2014]; Lavin [2013]; Thompson [2008 part 2]).

determination does not seem to fit into a metaphysics of events and states that efficiently cause each other. According to this metaphysics, everything that exists consists of events and states that stand in efficient causal relationships with each other. Hence, only an efficient causal explanation, not, say, a teleological explanation, can accurately describe what happens.⁵⁷ But if my activity consists in events and states that have efficient causes, it seems like my “intervening in the flow of events is just another part of that flow. So how can it count ... as an intervention—or, for that matter, as mine?” (Velleman [2000b] 5). More simply, I am not an event or a state. How then can my action fit into a world of events and states standing in efficient causal relations? Where in this world am I, and what in this world can I do? If I am not in this world and my action is not of this world, how can I or it be anything?

There are three ways to go here. You might reject the presupposed metaphysics. You then must account for how human action fits into a different general metaphysics of reality. In contrast, you might stick with that metaphysics, claim that human action conflicts with it, and conclude that human action is a myth. Or, with the causal theorists, you might argue that the presupposed metaphysics and human action are compatible because the “role assigned to the agent by common sense reduces to ... causal relations among events and states” (Velleman [1992] 130). Their task is to answer the following question. Even if there are “mental states and events in abundance ... connected, both to one another and to external behavior, by robust causal relations, ... how [does] the existence of relations of these items ... amount to a person’s causing something rather than merely to something’s happening in him, albeit

57. I say ‘as concerns what happens’ because these restrictions need not apply to, e.g., abstract objects or mathematical explanations in order for the causal theory to be important and controversial. Since Velleman’s argument for the causal theory is not a list of states and events standing in efficient causal relations to each other, he must not think that all legitimate explanations are efficient causal explanations. He must instead think that only explanations of things *as happenings in the world* must have this kind of explanation.

something mental?” (ibid. 131) The task is to show that “agent (or, self-) determination consists in some ... causal structure ... of ... psychological functioning, ... characterize[d] without presupposing the very idea of agent determination ..., such that agent determination of action consists in such functioning” (Bratman [2001] 91-2).

Causal theorists grant that human action is unique, different from the things that merely happen to an agent and from other activities like digestion and the action of the other animals. They claim, though, that it does not require unique kinds of things in order to exist. It only requires a unique arrangement of the same kinds of things that make up other kinds of happenings. Causal theorists succeed if they can show that a human being acts if and only if and because some specific efficient causal structure of mental states and events exists. Human action would then be compatible with the presupposed metaphysics of events, states, and efficient causal relations despite appearances of conflict. Human action would then be a “process composed of simpler processes in which events are caused by other events [and mental states]. I can make something happen even though it is caused by other events [and mental states] ... because their role in its production can ... amount to mine” (Velleman [2000b] 6).

I shall argue that causal theorists underestimate the difficulty of this metaphysical reduction. They consistently violate the explanatory commitments of such a reduction by explicitly and implicitly referring to an unreduced notion of self-conscious activity through their explanations of the constitution of our action. They can eliminate some of these references. I will argue, though, that the metaphysical reduction of human action is in principle impossible. Every version of the causal theory must appeal to a means-end belief, and I will show that means-end beliefs invoke an unreduced notion of what I can self-consciously do. Hold to the presupposed metaphysics, then, and you must offer an error theory of

human action. Hold to the reality of human action and you must offer a different view of the world and the place of human action within it. Since to offer an error theory is to perform a human action, the first option is self-undermining. Hence, the metaphysics must go.

2. THE STRICTURES OF METAPHYSICAL EXPLANATION

Before I criticize the causal theory, let me say more about the success conditions of a metaphysical reduction and distinguish it from a conceptual reduction. I should note that causal theorists are rarely explicit about the explanatory ambitions and demands of the view. I take them at their word that they mean to reduce our action, and I here spell out the conditions on such a reduction. As far as I can tell, causal theorists must offer this kind of explanation for their view to have the consequences that they claim for it. As far as I can tell, these conditions on this view are essential to such an explanation. To a large extent, though, this section is a kind of rational reconstruction, not exegesis.

Whereas a conceptual reduction must show that anyone who understands two concepts must recognize the identity between them, a metaphysical reduction can allow for non-transparent identities in ways that are familiar from cases of non-trivial identities. For example, even though Mark Twain is Samuel Clemens, I might believe that Mark Twain is a hack while failing to believe that Samuel Clemens is a hack. So long as I am ignorant of the identity, I can believe both things even though he cannot be and not be a hack (in the same respect at the same time). With non-trivial identities, terms or phrases that refer to the same thing are not necessarily intersubstitutable. In the same way, even if water metaphysically reduces to H₂O, I might believe that water is refreshing while failing to believe that H₂O is refreshing. So long as I am ignorant of the fact that water metaphysically reduces to H₂O, I can believe

both things even though it cannot be and not be refreshing (in the same respect at the same time). In contrast, if it is a conceptual truth that a bachelor is a unmarried male, anyone who understands the concepts cannot believe that someone is a bachelor and is married or that someone is a bachelor and is female. Likewise, if it is a conceptual truth that a vixen is a female fox, anyone who understands the concepts cannot believe that an animal is a vixen and is male or that an animal is a vixen but is not a fox. In both cases, the set of concepts on each side of the 'is' of identity are intersubstitutable with each other because they have the same meaning.⁵⁸ In contrast, in a metaphysical reduction, terms or phrases that refer to the thing to be reduced are not intersubstitutable with terms or phrases that refer to the reduction base.

Causal theorists are metaphysical reductionists, not conceptual reductionists. Hence, they need not show that anyone who possesses the concepts of human action and whatever concepts figure in their account of its constitution will recognize the claimed identity. Nor need they show that any sentences that use concepts that refer to human action are intersubstitutable with sentences that use concepts that refer to its constitution. The success conditions of a metaphysical reduction instead in part depend on whether the account of the constitution of the thing can explain the essential features of the thing. The metaphysical reduction of water to H_2O must explain, say, the fact that salt dissolves in water in terms of the chemical properties of H_2O and $NaCl$ and general facts about chemical bonds. It must explain the fact that water evaporates in terms of the properties of H_2O and facts about molecular motion under relevant conditions. It need not explain the fact that water is my favorite clear liquid or the fact that I

58. I am being loose here since this kind of conceptual identity is not a conceptual reduction. I do not know of any uncontroversial conceptual reductions, but it will not matter for the purposes of this paper. The important part is what metaphysical reductions do not require and what they do.

prefer water in its liquid state to water in its solid state or gaseous state. A metaphysical reduction of human action to its constitution must likewise explain the essential features of action in terms of its constitution.⁵⁹

Although a metaphysical reduction is not a conceptual reduction, certain issues about concepts feature in the explanatory strictures of a metaphysical reduction. In order to explain something in terms of its constitution, the account of its constitution must explain the essential features of thing without using concepts that implicitly or explicitly refer to the thing in question. Otherwise, the constitution is explanatorily posterior to the thing in question, and a metaphysical reduction of the thing to its constitution is impossible. Whether water metaphysically reduces to H₂O then depends on whether an account of H₂O that does not use the concept water or any concepts that depend for their meaning on water can explain its essential features. If, say, the only way to account for the bonds between the hydrogen atoms and the oxygen atom is to say they are *watery bonds*, such a reduction is impossible. It is likewise impossible if the proper explanation of the nature of hydrogen or oxygen includes claims that they are the waterbuilding elements. In the same way, causal theorists must explain the essential features

59. Let me issue a warning. Metaphysicians currently are all about grounding, but I shall not discuss it. Two parts of this literature are important for present purposes. First, metaphysicians disagree about the relationship between the grounding relation and metaphysical explanation. Some think that they always come together, whereas others think that there is more to explanation than simply charting an independent metaphysical structure. Second, they disagree about the relationship between grounding claims and claims about metaphysical reduction. Some think that to claim that A grounds B just is to claim that B metaphysically reduces to A, whereas others think that metaphysical reduction requires something else in addition to grounding. Since metaphysical reductions are a type of explanation, your view on the first issue affects your view on the second issue.

While I think that recent interest in grounding is good for metaphysics, I doubt that it has consequences for my topic or, for that matter, for much of interest in practical philosophy. As far as I can tell, the same issues, criticisms, and defenses pop up once a position is stated in terms of grounding as it did when stated in other terms. My claim here is that causal theorists commit themselves to a specific explanatory project that they cannot pull off. My criticism holds regardless of whether this metaphysical explanation is the same as claiming that the constitution of an action grounds the action, whether it involves a claim about a grounding relationship and something else, or whether claims about grounding have no place here. If we must talk about grounding, it only changes the way that we talk, not what must be said or why we must say it.

of action in terms of an account of its constitution that does not use the concept of human action or any other concepts that depend for their meaning on it. In particular, since their goal is to explain the self-consciousness of human action in terms of its constitution, their account of that constitution cannot include self-conscious activities of any kind. They can appeal only to mental states, events, and efficiently causal relations. These mental states, events, and relations must be intelligible independent of any concepts of human action or self-conscious activity. The specific combination of them must explain the salient and distinctive features of human action.

Try it this way. Consider J. David Velleman's claim that the "problem of agency is thus independent of, though indeed parallel to, the mind-body problem. Just as the mind-body problem is that of finding a mind at work amid the workings of the body, so the problem of agency is that of finding an agent at work amid the workings of the mind" (Velleman [1992] 131). Strictly speaking, he is wrong about both the mind-body problem and the problem of agency. He in effect defines both such that only a reductive answer to the question is possible. An argument against non-reductive views by such definitional fiat is no argument at all. Still, go with him. Just as you cannot 'find a mind at work amid the workings of the body' if you invoke mental predicates in describing the relevant aspects of the body, so you cannot find 'an agent at work amid the workings of the mind' if you invoke agential predicates in describing the relevant aspects of the mind. Just think about if you did. In specifying the nature of the mental states, the events, and the relations, you would make ineliminable reference to human action. In order to understand them, you need an independent understanding of our action. With that understanding in hand, what work can the causal theory do? The whole would instead be metaphysically and explanatorily prior to the parts, and the parts would be understood with reference to the whole.

Such an account of the nature of the parts thereby cannot explain the whole. A causal theory of human action is impossible, then, if the proper account of the mental states, events, and relations refers to our action or self-conscious activity. Put the other way around, a causal theory of human action is possible only if a proper account of its parts does not reference the whole and this account explains the features of the whole.

My criticism of the causal theory is that this kind of independence of parts from the whole is unintelligible. Whenever a version of the causal theory seems to explain human action in terms of its constitution, it in fact appeals to an unreduced idea of human action or self-conscious activity. Whenever a version does not appeal to an unreduced idea of human action or self-conscious activity, it fails to explain human action. The causal theory cannot meet its explanatory burdens.

3. CAUSATION AND ADDITION

Donald Davidson's 'Actions, Reasons, and Causes' is undoubtedly the first cause of the causal theory of action in contemporary philosophy. It is only a slight exaggeration to say that most action theorists take this paper to establish the truth of the causal theory *in general* and leave open to dispute only issues about the proper way to spell out details of the causal theory. Since I shall argue that the causal theory is in principle mistaken independent of the details of particular versions of it, let me start things by presenting his argument for the causal theory.⁶⁰

60. Whether Davidson means to offer a reductive causal theory is unclear. On the one hand, his worry about responding to the problem of deviant causal chains by saying that the mental states must efficiently cause the bodily movements 'in the right way' makes it seem like shares explanatory ambitions with reductive causal theorists, as I explain in a bit. On the other hand, he explicitly invokes the idea of an action when specifying the intentional content of the mental states that constitute a practical thought without mentioning it. On my view, these two postures are incompatible. Either there is no problem with ruling out deviant causal chains by qualifying the description of the efficient causal relation in a way that invokes the idea of an action or there is a problem with invoking the idea of an action within the intentional content of the mental states that efficiently cause the bodily movements.

Davidson does not start his investigation from the first-person perspective that is essential to action. He instead starts from how a spectator might describe my action to a third person, from which he infers the metaphysical constitution of action. According to him, an action explanation “rationalizes” what happens by citing “something the agent saw, or thought he saw, in his acting” (Davidson [1963] 3). For example, when I reach on the bookshelf, you understand me as acting only if an answer to ‘What is JDF doing?’ makes my moving in this way intelligible. Say someone else tells you that I am looking for my copy of *Authority and Estrangement* in order to commit it to the flames. It entirely lacks abstract reasoning about quantity or number. It contains no experimental reasoning about matter of fact and existence. Hence, it contains nothing but sophistry and illusion. It must go. This account explains what I am doing in terms of why I am doing it. In doing so, it unifies my movements into a coherent process with respect to a goal. I am moving around *in order to* find the book. I am looking for the book *in order to* commit it into the flames. I am committing it to the flames *in order to* rid my library of volumes of divinity and school metaphysics. So, an action explanation explains what I do in terms of my conception of why I am doing it. It relates my doing to my thinking.

Causal theorists, though, need to explain the way that this thinking and doing consist of mental states and events. Otherwise, all that an action explanation says is that I act because of my sense of why I act. Davidson’s focus is on the thinking. His view is that a practical thought “consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that” this action “has that property” (ibid. 5). The ‘pro-attitude’ explains the goodness that I find in the action, whereas the belief explains why I take acting in this way to be a way to achieve, realize, promote, or whatever that standard of goodness. In this way, Davidson accounts for the agent’s practical thought in terms of a set of mental

states that constitute it. To take the previous example, the proper way to explain the action is not to link the phases of my action one to another with an 'in order to' teleological explanation. The proper way to explain it is instead to connect what happens with a belief and a desire of mine whose intentional contents stand in an appropriate rational relationship to each other and a description of what happens as an action. I am moving about the room because I desire to commit Moran, or his work at least, to the flames and I believe that I can do so only if I find my copy somewhere in my library.

An account of human action needs to include more than an account of the mental states that constitute my practical thought. Action involves more than thinking, except, of course, where the action is mental in nature. I might have a set of relevant mental states without 'acting from' them. For instance, presuming that a belief-desire pair constitute a practical thought, I might desire your attention and believe that I can get it by moving my arm. Yet I might do nothing. Even if you attend to me because my arm moves, I do not act if, say, a taser makes it move (unless I apply the taser to myself or get someone to do it, I suppose). In fact, even if I get your attention by moving my arm, I might get your attention only accidentally while I act from some other practical thought. I might be hailing the guy selling roses in order to take one home to my sweetie like all those swell guys do in the talkies. But then what is the relationship between the mental states and my bodily movements when an action explanation in terms of those mental states is accurate?

Davidson's response is that "the relation between [those mental states] and an action ... is ... that the agent performed the action *because* he had" those mental states, where the causality of 'because' is efficient causation (ibid. 9).⁶¹ Hence, the set of mental states, not the agent, efficiently causes the bodily

61. Although nothing in this paper turns on this issue, I follow most causal theorists in treating the bodily movement as the intentional action rather than the efficient causal complex of mental states and bodily movements. I in fact think that the best

movement. I have a practical thought when I have tokens of the specified types of mental states that have appropriately related intentional contents. I act from that practical thought when that set of mental states efficiently causes my movements. I act intentionally, then, just when and because a specified set of mental states efficiently causes relevant bodily movements. That is the link between thought and action that differentiates my intentional actions from everything else that happens my life.⁶²

Although few causal theorists accept the details of this explanation, they agree with Davidson about its form. For them, an account of human action has an additive structure of the form ' $x + y = \text{human action}$ '. 'X' is the placeholder for the relevant set of mental states, 'y' for the relevant bodily movements, and '+' for the relevant efficient causal relation between the mental states and bodily movements. Their main intramural disagreement is about which mental states are part of the relevant set that distinguishes a human action from everything else. This debate tends to consist of counterexamples wherein an event that is not a human action meets the details of a specific additive theory. Still, while the details differ, the general form of the explanation is the same. It is this form that I shall criticize.

4. DAVIDSON AND DEVIANCE

The elements that go into the causal theory must be intelligible independent of their role in constituting our action. The account of the mental states, the efficient causal relation, and the bodily movements must not make explicit or tacit reference to self-conscious action. Otherwise, the causal theory invokes

version of the causal theory goes the other way, but explaining why turns on metaphysical issues about cause and essence that are irrelevant to my interests in this paper.

62. The mental states are the proximate efficient cause of the movements, not just *an* efficient cause of them. An account of relevancy must pick out the proximate efficient cause from the chain of efficient causes through time and from the totality of efficient causes and background and enabling conditions at a time. For this paper, the details of that account do not matter. Likewise, my criticism does not depend on any particular metaphysics of causation.

what it means to explain, thereby violating its explanatory burdens. Violations might happen at any of the three points of the explanation. Although I think that in the end violations happen at all of them in a true account of our action, I shall focus my criticism on violations in the account of the mental states. In order to clearly present the kind of criticism I shall pursue, though, I in this section present the problem of deviant causation for the causal theory. While causal theorists acknowledge this problem and discuss it, I think that they do not always appreciate the *kind* of explanatory challenges that it presents. Getting clearer on it reveals a general explanatory issue with the causal theory.

Davidson himself presents this problem with the following case:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening hold on the rope he could rid himself of the weight and danger. The belief and want might so unnerve him as to cause him to loosen his hold, and yet ... he [does not] do it intentionally. ... [S]ince there may be wayward causal chains, we cannot say that if attitudes that would rationalize x cause an agent to do x , then he does x intentionally.

(Davidson [1973] 79)

A desire and a belief here combine to efficiently cause bodily movements in such a way as to satisfy that desire. Yet the climber does not act. As we might put it, what happens with him that leads to his partner's death happens 'in spite of himself'. The causal theorist thus must distinguish the cases when a belief-desire pair efficiently causing bodily movements constitutes an action from cases when it does not. Otherwise, the account incorrectly counts certain events as actions. At least, one way to understand the problem of deviant causation is as a difficulty about extensional adequacy. Davidson seems to take it this way. He thinks that the problem is that

the action on the one hand, and the belief-desire pair which give the reason on the other, must be related in two very different ways to yield [an action] explanation. First, there must be a logical relation. Beliefs and desires have a content, and these contents must be such as to imply that there is something valuable or desirable about the action. ... Second, the reasons an agent

has for acting must, if they are to explain the action, be the reasons on which he acted; the reasons must have played a *causal* role in the occurrence of the action. These two conditions on reason explanations are both necessary, but they are not sufficient, since some causal relations between belief-desire pairs and actions do not give [action] explanations.

(Davidson [1982] 173)

Causal theorists tend to take up the challenge in the last sentence of this passage, adding to the additive account in ways meant to improve extensional adequacy. They try to specify the kind of causal relationship or the set of mental states in such a way as to rule out deviant cases. Of course, these routes of response are not mutually exclusive. A response that unifies both of them is likely the best course.

I shall not take up these responses, though, because my interest is in a different way to understand the issue behind the problem of deviant causation. After all, extensional adequacy is easy to achieve. Just say that the mental states must efficiently cause the bodily movements “in the right way” and you have extensional adequacy (Davidson [1974] 232). Similarly, just say that the mental states must efficiently cause the bodily movements ‘in the way needed in order for there to be an intentional action’ and you have extensional adequacy. Neither of these accounts are informative, though, and they thereby do not count as an explanation. More to the problem for the causal theorists, neither of them can be part of a metaphysical reduction of our action. The second overtly refers to intentional action. The former is at best a placeholder for the account of efficient causation. Without that account, it tacitly refers to intentional action since ‘in the right way’ just means ‘in the way needed in order for there to be an intentional action’.

The problem of deviant causation is thus not merely about extensional adequacy. Causal theorists need an extensionally and explanatorily adequate account of our action. I shall argue that they cannot offer it because they cannot meet the explanatory requirements of the account. I will not focus on

the problem of deviant causation, though, nor will I consider responses to it. It is an instance of a more general explanatory challenge for causal theorists. Not only do they need to account for the efficient causal relationship, but they need to account for the mental states as well. They must not invoke intentional action in their accounts of both the efficient causal relation and the mental states. I shall argue that they cannot meet the extensional and explanatory goals with respect to the account of mental states. Any set of mental states that is extensionally adequate appeals to the idea of intentional action in one way or another. Any set that does not is extensionally inadequate.

5. CONSTITUTION AND CIRCULARITY

Most of the intramural discussion between causal theorists is about whether a specific set of mental states is sufficient for an account of intentional action. For example, Davidson's account comes in for criticism from almost everyone, Michael Smith excepted. As far as I know, though, no one disagrees with him about including a means-end belief. I shall criticize causal theorists on the grounds that they cannot invoke such a belief given their explanatory ambitions. Let me start with more familiar complaints, though, and show that they embody the kind of criticism behind the problem of deviant causation.

Start with the belief-desire view. One standard objection is that it leaves the activity of the agent out of the picture. For example, an unconscious desire and a belief might efficient cause bodily movements that might not seem like human actions. At least, these events might not seem to exemplify the aspects of our action that distinguish it from the action of the other animals.⁶³ The account thus fails to capture the self-conscious self-determination distinctive of human action.

63. See Velleman [2000b] 8-9 for a version of this argument that uses Freudian unconscious desires. See Velleman [1992] 126-7 for a version that uses unconscious desires and beliefs without the Freudian backstory.

Harry Frankfurt argues to this conclusion on the basis of a case distinguishing a willing drug addict from an unwilling addict. You and I might desire the kind of pleasure that only heroin can provide, at least given our habits. We each might know how to get the drug and what to do with it. We each might actually get the drug and use it, you happily and me despite much anguish and effort to resist it. You seem to be acting wholehearted. Your will seems to be your own. I, on the other hand, seem to act in spite of myself, to be overwhelmed by my cravings. Frankfurt concludes that whereas you are self-determining, “the force moving [me] to take the drug is a force other than [my] own” (Frankfurt [1971] 18). Of course, he does not mean that I am moved by something outside of myself in the way that an avalanche might move me from the top of the mountain to the bottom. What moves me is my addiction, and in that sense I move myself. In another sense, though, *I* do not move myself because I am not acting self-consciously. I am not acting as I want to act. In fact, I am acting precisely as I do not want to act, unlike you.⁶⁴

On this basis, Frankfurt concludes that a belief-desire account of the mental states that

64. Whether Frankfurt is a causal theorist is a bit of tough nut. Causal theorists like Velleman and Bratman take him as one of their own even though he criticizes what he calls ‘causal theories of the nature of action’ in “The Problem of Action”. Let me say something here about this criticism and why it does not show that he is not a causal theorist.

Frankfurt criticizes causal theories because he thinks that they posit only efficient causes that are *antecedent* to the movements that constitute an action, thereby missing how an agent relates to her action *as it is ongoing*. His argument for this conclusion, though, is invalid. He recognizes that causal theorists think that “the causal sequences producing actions are necessarily of a different type than those producing happenings, but that the effects produced by sequences of the two types are inherently indistinguishable” (Frankfurt [1978] 69). However, he thinks that this idea entails that it “is integral to the causal approach to regard actions and mere happenings as being differentiated by nothing that exists or that is going on at the time those events occur” (ibid. 70). That is to say that he thinks that all actions are at a temporal distance from whatever makes them actions because movements that constitute actions are intrinsically indistinguishable from movements that do not. Frankfurt is correct that these movements are intrinsically indistinguishable from each other. Because efficient causation is between distinct existences, what makes a movement an action is extrinsic to the nature of the movement. Still, this fact does not show that what makes a movement an action is at a temporal remove from the movement itself. A causal theorist might instead insist that the mental states that constitute a practical thought must continuously efficiently cause the movements for their entire duration, sustaining them for as long as they happen. In fact, Frankfurt’s view in that essay seems like just such a version of the causal theory. His criticism is thus only of what we might call ‘antecedent causal theories’, and the target of his criticism is the antecedent bit, not the causal bit.

constitute a practical thought is incorrect. As with the problem of deviant causation, the basic problem seems to be an issue about extensional adequacy. Most action theorists take it up that way, and they do Frankfurt dirty in the same way that he does Davidson dirty. I shall argue that the extensional issue runs in parallel with an explanatory issue in a moment. Let me follow the intramural debate for another step or two, though, before I make that point.

Frankfurt amends the belief-desire model with a second-order desire that takes a first-order desire as its object. A desire for an apple is a first-order desire, a desire to desire an apple second-order. Frankfurt does not invoke just any second-order desire in his account. He instead thinks that a human being acts self-consciously when the first-order desire that she desires to determine her to act in fact determines her to act in a way that she believes will let her achieve the object of that desire. More simply, I act self-consciously when the desire that I desire to be effective is in fact effective. I, unwilling addict, thus do not act self-consciously because I do not desire for my desire for heroin to determine me to act. You, willing addict, do because you do. Human beings are distinct from the other animals, at least with respect to our practical capacities, because we can have second-order desires about the efficacy of first-order desires.

One way to challenge Frankfurt is to ask why the fact that a desire is *second-order* makes a difference. After all, the fact that the unwilling addict can be alienated from her desire for the heroin does not seem to depend its being *first-order* desire. It seems to depend on its being a first-order *desire*. Alienation from second-order desires is likewise possible. Say you and I are gay men of the age of reason who grew up in the same intolerant religious environment. We each might desire intimacy with other men. We each might also desire to neither have nor act on the first-order desire. We might blush, feel

shame, and chastise ourselves in response to sexual arousal. We might have thoughts about going to hell and darker thoughts about how we might get there in bad moments. We might desire to never act on the first-order homosexual desires. After all, acculturation and indoctrination shape not only what we desire but how we relate to our desires. Still, you might identify with these thoughts, thinking the shame and all that jazz justified, and desiring that the second-order desire govern your first-order desires. I might come to know that religion is bunk and its psychological effects are bullshit. When I find myself blushing and ashamed, I might know that I ought not to be and ought to rid myself of whatever psychological states efficiently cause this reaction. I thus might have a third-order desire that takes the second-order desire as its object and stands against its efficacy in just the way that the second-order desire takes the first-order desires as its object and stands against its efficacy.

To posit a set of third-order desires in response to this criticism is to miss the point. It might be hard to think up counterexamples given the dizzying psychological hierarchy. The counterexamples, though, are not the source of the problem. They just make it vivid. The problem is a formal feature of desires. I can take a critical stance with respect to any of them, and I can be alienated from any of them, regardless of their intentional content.⁶⁵ A problem with an account in terms of first-order desires and a means-end belief is thus a problem with an account in terms of N^{th} -order desires and a means-end belief. As Gary Watson puts it, since “second-order volitions are themselves simply desires, to add them to the context ... is just to increase the number of contenders; it is not to give a special place to any of those in contention. The agent may not care which of the second-order desires win out. The same possibility arises at each higher order” (Watson [1975] 18).

65. As I shall explain in a bit, Velleman in fact denies this claim. I am using this account in order to get the kind of criticism in view, and nothing that I say will turn this claim.

If not desires, then what? Specifying that the mental states are ‘of the right kind’ is as inadequate as specifying that they must efficiently cause the bodily movements ‘in the right way’. It is extensionally adequate but explanatorily empty. Moreover, without further explanation, ‘of the right kind’, like ‘in the right way’, is incompatible with a metaphysical reduction of our action. Either explicitly or tacitly, saying that the mental states are ‘of the right kind’ means that they are the set of mental states needed in order for bodily movements that they efficiently cause in the right way to be an intentional action. Although explanatorily adequate, this characterization of the mental states invokes the idea of self-conscious action and thereby violates the explanatory strictures of a metaphysical reduction.

Although not as overtly, many causal theorists violate these explanatory strictures in response to this kind of problem in two different ways. They either invoke an *agential act*, or they invoke a kind of mental state that appeals to the idea of self-conscious activity. For the first kind of violation, take Frankfurt’s claim that when “a person identifies himself decisively with one of his first-order desires, this commitment ‘resounds’ throughout the potentially endless array of higher orders. . . . The decisiveness of the commitment ... means that he has decided that no further question about his second-order volition, at any higher order, remains to be asked” (Frankfurt [1971] 21). While the idea of a person identifying himself might contribute to an extensionally adequate account of our action, it violates the explanatory strictures of the causal theory in the same way that invoking the idea that the mental states must be of the right kind. After all, identification is something that an agent self-consciously does, and self-conscious activity is the object of explanation for the causal theory. Likewise for notions of forming a mental state, endorsing it, and many others. To the extent that this stuff has a role to play in a metaphysical reduction of our action, it is to be explained, not to do the explaining.

The second kind of violation is subtler and more problematic. One way to respond to the problem about desires is to add other kinds of mental states to the set that constitutes a practical thought. In particular, causal theorists often add intentions to desires and means-end beliefs in order to correct the extension of their account. Intentions might play any number of roles in a theory. Maybe they help explain how different phases of an action unify into a single action rather than a number of discrete actions, as when I perform the action of hitting every button for every floor rather than the set of actions of hitting single buttons until I hit them all. Maybe they help explain self-determination, as intentions might seem to differ from desires because they seem more directly sensitive to and expressive of evaluation and choice. Whatever the reason why a particular theorist posits an intention, though, the standard account of them says that the intentional content is an action. I intend to ϕ —I intend to go to the store in order to get my sweetie some jelly, or I intend to go home in order to cook with him. After all, to intend is not merely to want a state of affairs to obtain. It is to commit myself to bring about some end by taking requisite means to it. I do not satisfy my intention if the end merely happens to come. I do not satisfy it even if I feature in the efficient causal history of the end in some way or other. I only satisfy it if I self-consciously act in the way that I intend to act and thereby bring about the end. The object of an intention exists only if I successfully intentionally act as I intend to act. The object of an intention is a self-conscious action. As Matthew Boyle and Douglas Lavin put it, “to represent my doing A is to represent, as it were, a kind of state of affairs whose obtaining is my having intentionally caused it to be” (Boyle and Lavin [2010] 173).

Whatever the proper account of intention and its place in human agency, it has no place in a metaphysical reduction of our action, at least not on this simple account of it. Causal theorists cannot

invoke it because it refers to the idea of a self-conscious action in the intentional content of an intention.

To go to Boyle and Lavin again,

if we cannot hope to give an account of wanting to do A that is independent of the idea of intentional action, then neither can we hope to give a reductive account of intentional action as a matter of movement caused in the right way by such wantings. An account of what it is to want to do A must rather presuppose an account of action, an account of the kind of event (S's doing A) whose coming to be is the subject's intentionally causing it to be. (ibid. 173-4)

Just so, if causal theorists cannot account for the nature of an intention except through reference to intentional action, they cannot metaphysically reduce our action. An account that invokes an intention might be extensionally adequate. It is as explanatorily empty, though, as an account of water that invokes the idea of water in its account of the elements that constitute it or their arrangement and bonds. As with intentions, so with desires, plans, policies, commitments, or any other kind of mental state a causal theorist might invoke.

6. SELF-REFERENCE AND INTENTIONAL CONTENTS

Causal theorists seem to need to invoke an intention or something like it in order to account for a practical thought. The simplest way to invoke it, though, is incompatible with a metaphysical reduction of our action because its object is an intentional action. They thus need to invoke an intention while specifying its intentional content in a different way. Can they account for the intentional content of a practical thought while not appealing to anything that the metaphysical reduction is to explain?

Perhaps. Although they do not put it to precisely this use, some causal theorists claim that the intentional content of an intention must be self-referential in order to capture the self-consciousness of our action. Velleman claims that an intention is a kind of self-verifying belief. This intentional content of

this belief is something like ‘I will ϕ because of this very belief’. It is not self-verifying because it has an internal tendency to become true. It is instead self-verifying because I have a desire for, say, self-knowledge that along with ordinary desires and means-end beliefs efficiently causes me to act in a way that makes this belief true and thereby satisfies the desire for self-knowledge. Keiran Setiya similarly claims that an intention is a desire-like belief whose intentional content is likewise something like ‘I will ϕ because of this very belief’. It is a belief because it represents something as the case, desire-like because it makes the case what it represents as the case. Hence, this self-referential belief has an internal tendency to become true whereas Velleman’s self-referential belief has an external tendency to become true.

The self-referentiality of these mental states might not seem to help. After all, they still refer to an intentional action—“the phrase ‘doing A [or, in my terminology, ϕ -ing]’ remains in the content clause, and this sort of content is one to which the causal theorist is not entitled to appeal” (Lavin and Boyle [2010] fn.196n19). However, causal theorists should deny that the content of an intention refers to intentional action as such and instead claim that it refers to *the constitution of an intentional action*. The intentional content then is something like ‘my body will move in such-and-such a way because of this very belief’ or, the other way around, ‘this very belief will efficiently cause my body to move in such-and-such a way’. A full development of this view must specify the other elements of the constitution of an intentional action, including the causal roles of desires and whatever other mental states partially constitute a practical thought. Still, it might seem to provide a model of how to respond to the challenge. It might seem to offer an extensionally adequate account of our action that is explanatorily adequate as well. After all, if an intention to ϕ offers extensional adequacy, why should an intention to do-what-constitutes-the-action, though not in those terms, do any worse? If a metaphysical reduction of action is

at least a live possibility, why would this account not be as explanatorily adequate as an account of water in terms of its constitution?

To pull such an explanation off requires a lot of work. Causal theorists must spell out the constitution of an intentional action in full, spell out the intentional contents of all the mental states that jointly constitute a practical thought without appealing to intentional action, and show that this account explains the features of intentional action. On first glance, though, nothing seems to rule it out. However, I shall here argue that this explanation cannot work. Every account of the mental states that constitute a practical thought must include a means-end belief, and a means-end belief just is a belief about what I can self-consciously do in order to achieve my end. Let me explain.

Take as the exemplar something like Velleman's account of a practical thought. Although his view changes in various ways through the years, the basic structure starts with normal desires for various objects and states of affairs and a desire with a special object. For present purposes, assume that it is a desire for self-knowledge. According to Velleman, this desire explains how an agent surveys her desires—or, more accurately, their objects, or the way that the world appears to someone from the standpoint of a desire—and chooses what to do from amongst them. Presume that an account of the efficient causal interaction between the desire for self-knowledge and the regular desires is possible. This efficient causal interaction in turn produces an intention—that is, a belief that I will move my body in some way for some purpose because of this very belief itself. I satisfy the desire for self-knowledge if this belief becomes true. After all, I cannot know that I move my body in this way because of this belief if I do not do so. This desire and the intention, along with the regular desire that combines with the desire for self-knowledge to produce the intention, thus combine to efficiently cause me to move in the way that the

intention specifies. If I actually do it, I have self-knowledge, thereby satisfying both the original desire and the desire for self-knowledge.

Whatever questions you have about this account, one thing that is clear is that it is inadequate as it stands because it cannot explain how the intention comes about because of the original desire and the desire for self-knowledge. I do not here mean to question the possibility of an efficient causal account of the interaction of mental states and how it somehow tracks the rational relationships between the intentional contents of those states. I instead mean to say that these two desires cannot explain the intentional content of that intention on their own even if such an efficient causal explanation of mental processes is possible. The desire is for an object or a state of affairs—let us say to have the one ring that rules them all in my possession. The desire for self-knowledge is a generic desire for my beliefs to be true and whatever else must be the case for them to constitute knowledge. The combination of these desires alone cannot explain how I form an intention to move my body some way. They can only explain this intention in combination with a means-end belief. Every casual theory, no matter what separates it from others, needs this kind of belief to explain how an agent moves from a generic practical thought about an end to a more determinate thought about an action that includes a means and a way to achieve the end

However, a means-end belief is incompatible with the possibility of a metaphysical reduction of our agency. A means-end belief is about the causal relationship between two things, but it is not *just* a belief about causal relationships.⁶⁶ Otherwise, my belief that the fire in the fireplace efficiently causes my cat's belly to warm would be a means-end belief. Nor is it merely a belief about causal relationships that

66. I am ignoring the difference between constitutive means and causal means for the purposes of exposition. Similar points apply to constitutive means. Not just any belief about a way for something about me to constitute something else is a belief about a constitutive means. Only beliefs about ways for me to self-consciously do something do.

in some sense involve me. Otherwise, my belief that I can get to China by quantum tunneling would be a means-end belief. Nor is it merely a belief about causal relationship where I am in some way the 'cause'. Otherwise, my belief that my beating heart circulates my blood would be a means-end belief. Nor is it merely a belief about a causal relationship where I in some sense as agent am in some way the 'cause'. Otherwise, my belief that my talking causes perturbances in the air is a means-end belief even if I know that causing perturbances in the air is neither an end of mine nor a means to any end of mine. A means-end belief is essentially about what I can self-consciously do in order to achieve some end. Only then does it capture the way that practical thought is about a way *for me to act*.

For me to have a means-end belief, then, is for me to have a practical thought about what I can self-consciously do in order to pull off my end. Moreover, the presentation of the intentional content of this belief under this guise is not an accident. This thought is not merely of a way that *someone* can pull off an end of this kind. It represents a way that I take to be open to *me* to pull off my end. A thought of a means that is not a means for me is, in the sense relevant to a practical thought, no thought of a means at all. Moreover, the thought about a means for me is not a thought about someone who is in fact me. It is a first-personal thought. After all, I cannot have a means-end belief about something that I *in fact* can do in order to pull off my end if I do not *think* that I can pull off my end in that way. Skills and opportunities of which I am unaware cannot feature in my means-ends beliefs. To represent a means is to represent an act type as something that I can do in the distinctively self-conscious way in order to pull off my end.

The only option for the causal theorist is to construe all the beliefs that are part of the reduction base in a way that does not make reference to means. They must only be about causal powers, and these causal powers must not be powers of *mine* by which I can realize a practical thought. They must just pick

out physical possibilities. It is not clear that this kind of thing is possible, but say that it is. How could the combination of this kind of belief and a desire 'stand in for the agent'? They cannot. To understand the problem, consider an issue that Davidson raises for Paul Churchland. Davidson points out that the agent might have a relevant desire and belief but not recognize the logical connection between their intentional contents and hence be unable to acting for the practical thought that they could constitute.⁶⁷ Similarly, an agent who recognizes that a causal route to achieving an end is open *to someone* does not recognize a possible means unless she recognizes that she herself is that someone. Moreover, she does not recognize a possible means unless she recognizes that the causal route is open to her to take in the sense of the kind of thing that she can self-consciously do. After all, I can only act from a recognition of a causal pathway that can help me achieve my end if I take that causal pathway to be either the kind of thing that I can do or the kind of thing that I can control by doing (or avoiding doing) other things. Yet then a characterization of the mental states that constitute a practical thought ineliminably refer self-conscious action in a way that undermines the possibility of a metaphysical reduction of action. A metaphysical reduction of our action is impossible.

7. The Way Forward

Think back to Velleman's claim that the problem of agency and the mind-body problem have a similar structure. He formulates the problem in such a way that only a reductionist about mind and agency counts as addressing the problems. Yet the problems do not dissolve once reduction fails. This failure instead shows that Velleman did not formulate the problems correctly.

67. Davidson uses 'reason' or 'primary reason' where I use 'practical thought'. He also says that an agent with these mental states who does not recognize the logical connection between their intentional contents 'has' the reason but cannot 'act for it' where I say that the agent 'could have' the practical thought but does not 'have' it. Nothing depends on these differences.

The more basic challenge of action theory is to show that human beings act in the same world in which we digest, in which the others animals act, digest, and live their lives, in which plants grow, and in which stuff happens to non-living things. Velleman and other reductionists twist this challenge into their own version, but the more general challenge is the real source of interest in action theory. Reductive causal theorists start with a metaphysics appropriate to the non-living things and try to apply it to our action, in effect ignoring everything else in our lives, the lives of the other animals, and the lives of plants. They thereby miss essential features of our action.

A different view does not presuppose that there is one metaphysics to rule them all. It instead starts with an account of the basic difference between the living and the non-living. Some of what happens in the lives of living beings are their doings in any of the various ways that living beings do things. A general metaphysics of activity and the difference between it and mere happenings is beyond my scope here. Imagine, though, that we have it in hand. Our action is one of a variety of activities of living beings, fitting into the same general metaphysics of activity but different than other activities because it is self-conscious. Such an explanation of our action fits it into a general account of the world and shows that we can act in the same world in which everything else happens. Yet it does not try to reduce our action to something else. It is a non-reductive account of our action that is not silent on the basic challenge of action theory. That is the way to go in action theory.

THE SUBSTANCE OF CONSTITUTIVISM

The normative requirements in the theory of motivation do not merely describe externally observable (or internally *observable*) patterns; they are internalized, they govern the agent's critical faculty, they characterize him as the source of his actions and thoughts. He does not choose them, for choices must issue from him if they are to be his, and this means that they must be the product of the determining principles which constitute him as the source of his choice, and which could not be chosen by him because in their absence there would be no *he* to choose. ... There is nothing regrettable about finding oneself, in the last analysis, left with something which one cannot choose to accept or reject. What one is left with is probably just oneself

(Nagel [1970] 22-3)

1. THE PRACTICE OF REASON

ETHICS is a crusade against solipsism. The ethical life is a constant struggle to act decently and stand in decent relationships despite natural, personal, psychological, political, social, and historical obstacles and threats, among others. You are in an important way just like me and must figure in my practical thought in much the same way that I must. Your good is to condition how I act in much the same way that mine is, even if it massively affects my life. This impact might be all at once. At the limit, your good can require me to sacrifice my good and perhaps even my life. Your good, though, should influence me in mundane ways all of the time. Telling the truth can hurt you, can hurt me, and yet be necessary all the same. Being there for you can get in the way of my pursuits and still be needed. Being decent can take all of the effort in the world if indecency is an everyday and unconscious part of social life. Life is difficult, and this difficulty ends only when life does. Still, we must be decent to each other.

Yet the struggle of ethics is not just a lived part of life. It is also a basic struggle of practical philosophy. Why are you to figure in my practical thought in much the way that I am? It is my life, after all, not yours, and I can only live my life, not yours as well. Are we subject to moral requirements in our action as such? If so, what explains why we are? My aim in this paper is not to answer these questions. I

instead mean to investigate a view about the nature of normativity for which I argue in “The House of Goodness”, called *constitutivism*, and explain what it implies about the structure of an answer to these questions.

Constitutivism is a metaphysical view about the nature of normativity that says, very roughly, that what something is determines what it ought to be, at least with respect to things by nature subject to normative standards. As I prefer to put the view, its basis is the view that a single principle describes the nature of a capacity and is thereby normative for its development and exercise. By this point, philosophers bearing the banners of the houses of Aristotle, Hume, Kant, kinda-Kant, and Nietzsche explicitly endorse the label.⁶⁸ Whereas they disagree over which normative standards govern our action as such, they agree that we are subject to such a standard only if the nature of action establishes it. Their disagreements over normative matters are at bottom disagreements over the metaphysics of our agency. They agree that to know whether and why you count for me in much the way that I count for me, we must understand our nature. To make progress in ethics, we must do metaphysics whose goal is self-knowledge.

My goal in this chapter is not exactly to offer an account of our action and its normative standards. Instead of defending a particular version of constitutivism over alternatives, I mean to

68. I argue elsewhere that pretty much everyone is a constitutivist because everyone thinks that the basic normative standards to which an agent is subject depends on the capacities they by nature possess given their kind. This claim is less dramatic than it sounds because constitutivism as such is just the claim that the functional nature of capacities is the basis of all normativity. It is distinct from particular constitutivist claims about the nature of the relevant capacities. Hence, intellectualist views of practical reason are compatible with constitutivism proper even though I argue that the view of practical reason is incorrect. Still, constitutivist label itself is increasingly popular. See, for example, Michael Thompson [2008 part 1] for one version of Aristotelian Constitutivism, Michael Smith [2013] for a Humean version, Christine M. Korsgaard [2009] for the paradigm Kantian version, J. David Velleman [2009] for the most recent and developed version of his self-named kinda-Kantian version, and Paul Katsafanas [2013] for the Nietzschean version. No doubt some enterprising soul will soon fill the deplorable Hegelian-shaped gap in this literature.

investigate its explanatory structure and epistemological commitments. Philosophers for and against it generally assume that the nature of our action establishes a normative standard for our actions only if that normative standard follows analytically from the concept of self-conscious or rational action. Certain normative standards in fact follow in this way from that concept. In particular, a version of the instrumental rule follows in this way, as I shall show.⁶⁹ It might then seem like other normative standard for our actions either likewise follow from the concept of self-conscious action or else they have only a secondary normative status with respect to our action. That is, it might seem like either they follow in the way the instrumental rule follows or it has a normative status that they lack.

I shall argue, though, that this view distorts the explanatory structure of constitutivism. The key to understanding why is placing the constitutivist view our action within a view of the nature of the capacities of the living in general. Ethicists tend to focus on our action without thinking about how it fits into a constitutivist view of the nature of normativity in general. Understanding human action is admittedly more pressing and interesting than understanding human digestion or feline action. Still, these different capacities embody the same metaphysical and explanatory structure. I shall argue that the claim that a standard governs our action as such only if it analytically follows from the concept of self-conscious action is as plausible as the claim that a standard governs our digestion as such only if it analytically follows from the concept of animal nutrition. Just as substantial aspects about the feline and human forms of animality determine what cats and humans should eat and how they should process it, so substantial aspects of the human form of self-conscious agency determines what we should do and why we should do it.

69. I say the instrumental *rule*, not the instrumental *principle*, in order to avoid confusion when I say that the instrumental *rule* is part of the *principle* of our will.

By doing more general metaphysics, then, we shall understand that constitutivism is more complex, flexible, and interesting than its supporters and detractors take it to be. I shall close by arguing that the route to knowledge of the nature of our agency and hence to the basic normative standard for our action as such is fundamentally ethical, not epistemological. It is the work of a good human life, with everything that involves, not the work of philosophy, or at least not the kind of analytic task that differs from the thinking, reflecting, and acting essential to a human life and philosophy as well. We can articulate the metaphysical and explanatory structure of our capacity to act through traditional philosophical means. To fill in the details of an account of the principle of that capacity, though, is to exercise it, to reflect and act in the way characteristic of a self-conscious agent. Ethics is a practical subject, a practical struggle.

2. FORMS OF REASON

Animals have wills. To be an animal is, in part, to be able to act, and a will is a capacity to act. As far as we know, though, only we human beings have practical reason. Like the wills of the other animals, it has an essential link with the ability to act. What, then, is the relationship between practical reason and the will? Immanuel Kant tell us that “the will is nothing other than practical reason” (Kant [1785] 4:412). Aristotle tell us that the conclusion of a practical syllogism is an action—not a judgment or a belief about how I should act, not a decision or an intention to act in some way, but the action itself. If practical reason is the will and if the will is our capacity to act, practical reason is our capacity to act. Practical reason is our kind of will. It thus differs from theoretical reason in kind, not merely in the contents of its representations. In this section and the next one, I shall reflect on this view and explain what it means for

practical reason to differ from theoretical reason in this way. Although this discussion is at some remove from questions about the proper way to develop constitutivism, it is essential to get it into view in order to understand various aspects of the argument to come.

One way to get into the view that practical reason is our kind of will is to consider the more familiar view that practical reason and the will are different capacities in us. If all animals have a will and only human beings have practical reason, perhaps we have practical reason in addition to having a will. Our will is the same as the wills of the other animals. It is our ability to act and explains aspects of our agency that we share with them. Practical reason accounts for the other aspects of our agency that distinguish us from them.

This view might seem like common sense, but I do not think that it captures the role of either practical reason or the will in our agency. While I cannot argue to that effect here, let me try to talk you into it. Think about instrumental thought. Whatever else it involves, an exercise of practical reason includes thought of the form 'do *this* in order to achieve *that*'. This thought need not run through my mind. It instead describes what I am doing in a way that picks out the process as an intentional action. The part-whole structure of the doing realizes the means-end structure of the thinking. This is why I can explain to you what I am doing by saying 'X-ing by Y-ing' or 'Y-ing in order to X', at least when I exercise my capacity to act well. In this way, 'Why are you Y-ing?' and 'How are you X-ing?' ask indifferently after the part-whole structure of the doing or the means-end structure of the thinking.

Yet the action of other animals likewise has a part-whole structure. Their thought thereby has a means-end structure. Not only can my cat run across the room, but he can run across the room in order to pounce on the stuffed mouse. He succeeds by running and pouncing. He fails if he gets blocked on the

way. Not only can he hide upon seeing me retrieve his carrying case, but he can do so by sprinting into the bedroom and diving under the bed. He succeeds by getting under the bed, out of sight and safe from capture. He fails if squished to the floor mid-spring and caged in his case. Of course, he cannot recognize that his thought has a means-end structure or that his action has a part-whole structure. He is not self-conscious. Still, although he lacks the concepts of a means, an end, a part, and a whole, the part-whole structure of his doing realizes the means-end structure of his thinking.

A mark of the exercise of practical reason in us, then, is present in exercises of the wills of other animals. How might we try to salvage the view that practical reason and the will are separate capacities? One way is to remove instrumental thought from the exercise of one of the capacities. To remove it from the exercise of the will, though, leaves our will with little to do in our agency. Its exercise is perhaps only a point-like decision whose object lacks internal structure. This hardly resembles the wills of other animals. It hardly is a capacity to act. Our action realizes the object of the will. If that object lacks structure, so does our action. No one ever does *this* in order to achieve *that* or does *that* by doing *this*. In losing the means-ends of the thinking, we lose the part-whole structure of the doing. Instrumental thought must be part of the exercise of the will.

To remove instrumental thought from the exercise of practical reason, though, likewise leaves practical reason with little to do in our agency. Its exercise is perhaps only a judgment whose object has no structure and hence cannot correspond to anything in the doing. After all, if anything in its representation corresponds to the part-whole structure of an action, it includes instrumental thought. What, though, does such a thought have to do with action? Whatever else is wrong with this idea, it hardly is a capacity for *practical* reason.

Instrumental thought, then, is a mark of the exercise of the will in the other animals and a mark of the exercise of practical reason in us. There are fancy ways to try to account for this fact within a view that separates practical reason and the will in us. I argue against this view in “Intellectual Isolation”. Here I mean to start with the thought that instrumental structure characterizes the exercise of practical reason in us and the exercise of the wills of the other animals because practical reason is our kind of will. On this view, while every animal has a will, not every will is like ours. Like others, practical reason is a capacity to act. Unlike them, it is a self-conscious capacity to act. Or, to make the point another way, ‘practical judgment’ is the name for our kind of decision, the kind characteristic of a self-conscious being. To judge practically is to self-consciously decide. Instrumental thought thus characterizes the will of every animal.

So say that practical reason is our kind of will and thus our capacity to act. The principle of this capacity, then, describes it and is thereby normative for its exercises. What is this capacity a potentiality to do and thus what is the content of the normative standard that governs our action as such? To answer these questions, we need an account of the function of practical reason. To this end, I will explain what it means for practical reason to be a distinct form of reason. I shall use some ideas from G.E.M. Anscombe’s *Intention* about the difference between theoretical and practical knowledge. While discussing knowledge might seem out of place in an account of reason, the metaphysics of capacities licenses it. Let me explain.

While all capacities are potentialities, not all potentialities are capacities. A mark of a capacity is that its actualizations divide into the good and bad, the successful and unsuccessful, the perfect and defective, the complete and incomplete. A capacity is, after all, a capacity to *do* something, and capacities differ from each other based on what they are potentialities to do. A capacity to walk differs from a

capacity to fly, say, given the differences between walking and flying. The nature of an activity thereby distinguishes a capacity to do that kind of thing from others. Only a being with the relevant capacity can engage in that kind of activity. In this way, capacities make possible their exercises, which are subject to normative standards that derive from the nature of the capacity. Yet they also make possible characteristic defects in their exercises. If I can walk, I can trip, slip, stumble, or tumble, and only a being who can walk can trip, slip, stumble, and tumble. Walking is doing what I have a capacity to do or succeeding in doing what I am doing. Tripping and slipping are deviations or imperfections in exercises of that capacity. They constitute failures of those exercises if they are severe enough. A walk to the store that ends at the store is a successful though imperfect exercise of my capacity to walk if it includes a few stumbles or other imperfections along the way. A walk to the store that ends with me flat on my face with a fractured ego and an even worse leg is a failed exercise of that capacity. In this way, a capacity brings with it two classes of its actualizations, the good and the bad.

A capacity has metaphysical and explanatory priority over its exercises in the sense that it makes them possible. However, one way to understand the nature of a capacity is through its perfect exercise. After all, a perfect exercise meets the internal normative standard for exercises of that capacity. Since the principle that is normative for those exercises also describes the nature of the capacity, to understand the perfections of the exercise is to understand the nature of the capacity. To focus on theoretical and practical knowledge is to attend to the perfect exercises of theoretical and practical reason in order to understand their nature. It is to understand what unites them as distinct *forms of reason* and differentiates them as *distinct* forms of reason. Their unity and diversity comes out in how they differently determine an abstract genus in terms of the way that thought relates to its object in exercises of them and in what

that entails about the nature of the thought and object. Here goes.

When I exercise theoretical reason, I represent objects that do not depend on that representation. I represent them as independent of that representation. A tree, say, is what it is regardless of my thought about it. It is an ‘independent existence’, and this aspect of my relation to the object also characterizes my representation of the object. That aspect is, as it were, a formal part of that representation, not part of its content. It is the way that I represent the object, and I can articulate it by reflecting on the exercise. Theoretical reason is a capacity to represent independent existences as such. I thereby must represent the world as it is in order to exercise theoretical reason well. To exercise it well, then, is to non-accidentally represent an independent existence accurately. Because the object is an independent existence, I represent it *accurately* when I represent it as it is independent of my representation of it. I *non-accidentally* represent it accurately when I stand in a proper relationship to it like, say, a perceptual or testimonial relationship. The basic question of theoretical reason is then ‘Whether *P*?’ or, generally, ‘What is the case?’ To answer it is to have “speculative knowledge” that “is derived from the objects known” (Anscombe [1957] §48).

Practical reason differs theoretical reason in those respects. When I exercise it, I represent objects that depend on my representation of them. I represent them as depending on that representation. For example, I represent cooking with my sweetie in order to spend time with him, and I spend time with him by cooking with him because of that representation. The object—the action—by nature depends on my representation of it in that way. That is why I am not acting if I sleepwalk through similar bodily movements. The object is a ‘dependent existence’, and this aspect of my relationship to the object characterizes my representation of the object. That aspect is, as it were, a formal part of that

representation. It is the way that I represent the object, and I can articulate it by reflecting on the exercise. Unlike the object of a theoretical representation, though, this aspect is also part of the content of a practical representation. I represent dependent existence, which is an object that depends on that very representation of it. I thereby must realize the object of my representation in order to exercise practical reason well. To exercise it well, then, is to non-accidentally represent a dependent existence accurately. I *non-accidentally* represent this object when I act in some way because doing so realizes this object. I non-accidentally represent this object *accurately* when the true description of my action as an intentional action is a determinate version of the content of my representation. The basic question of practical reason is then ‘Am I to ϕ ?’ or, generally, ‘How am I to act?’ To answer it is to have “practical knowledge” that is “the cause of what it understands” or ‘of objects that derive from this knowledge’ (ibid.).

As we might put the basic idea, a theoretical representation *describes* its object while a practical representation *produces* its object. Moreover, the object of a practical representation is per se the object of that representation whereas the object of a theoretical representation is not. The objects and the representations of them reflect these facts about the relationships between them. Theoretical and practical reason thus are different forms of reason because of the interdependent differences between the nature of their representations, their objects, and the relationship between representation and object.

One way to get into this account is to contrast it with the view that differentiates theoretical and practical reason in terms of propositional attitudes with different directions of fit. After all, philosophers often read the direction-of-fit view into Anscombe’s discussion of the different ways that a list can relate to the items in a shopping cart (ibid. §32). This discussion is part of her extended argument for

distinguishing practical knowledge from the “incorrigibly contemplative conception of knowledge” that dominates modern philosophy (ibid.). Hence, it might seem like it is the basis of an account of practical knowledge that reveals the nature of practical reason. The direction-of-fit view, though, is in fact diametrically opposed to the distinction between theoretical and practical knowledge as I and, in my opinion, Anscombe mean to draw it.

According to the direction-of-fit view, a theoretical representation consists of a mental state with mind-to-world direction of fit like a belief, and a practical representation consists of a mental state with world-to-mind direction of fit like a desire. The view says that these mental states share a single kind of relationship to an object, though the direction of the relationship flips. Since an object is independent of a belief about it, then, it must be independent of a desire about it in the relevant sense. Maybe a desire causes its object to exist whereas the existence of an object of belief causes the belief. Maybe instead mismatch between a desire and its object indicts the object whereas mismatch between a belief and its object indicts the belief. Regardless of whether a direction-of-fit theorist prefers a causal or normative version, or any other version for that matter, still, metaphysically, the object of a desire is not *essentially* an object of that desire. It cannot be a dependent existence because the object of the belief is an independent existence. This is, in effect, to deny the practicality of a practical representation as characterized by the essential interdependence of a practical representation and its object. Yet if you stick with that practicality, you cannot offer an account of a single relationship that theoretical and practical representations can bear to the same object in different directions. A representation of an independent existence as such does not have the same kind of relationship as a representation of a dependent existence as such except that it is ‘the other way around’. To represent an independent

existence as such is, after all, to represent it as independent of my representation of it. A theoretical representation thus lacks a relationship to its object that a practical representation possesses. To lack a relationship is not to possess it 'the other way around'. The possibility of practical knowledge is thereby incompatible with the direction-of-fit view.

These claims are about the perfect exercise of these capacities and thus about the nature of the capacity. They are not claims about every exercise. Defective exercises happen. I can get it wrong about the tree. I can fail to act as I practically represent. The metaphysics of capacities allows for this deviance. After all, the principle that describes the capacity is normative for its exercises, not descriptive of them. Likewise, I can exercise these capacities in derivative but not defective ways. I can ask not only 'What is the case?' but 'What would have been the case?', can think not only about real objects but also imaginary objects. I can ask not only 'How am I to act?' but also 'How would I have to act?', can think not only about actual acting but also about hypothetically acting. The metaphysics of capacities allows for these derivative exercises without saying that they are defective. 'What is the case?' and 'How am I to act?' are the *basic* questions of theoretical and practical reason, not the *only* questions that I can address in exercises of these capacities. Defective exercises are exercise that fail to answer those questions well. Derivative exercises are exercises that answer a different version of those question for a different purpose. Still, an account of the nature of the capacity focuses on the basic questions in order to understand the function of the capacity in our lives. Such an account must not stick to what the factual and counterfactual, actual and hypothetical, share. Otherwise, practical reason does not have an essential connection to action and is not our kind of will, nor does theoretical reason have an essential connection to knowing and understanding our world. They would thereby have little relationship to the wills and

understandings of the other animals.

3. REASON AND THE FIRST PERSON

Theoretical and practical reason are self-conscious capacities. This self-consciousness is important for their development and exercise. I shall start with a difference between the first-person in perfect exercises of them. I will then discuss the role of reflection in developing and exercising them.

Representations relate to their objects differently in exercises of theoretical and practical reason. In an exercise of theoretical reason, the object is independent of *my* theoretical representation of it. Some objects, like trees, are independent of anyone's representation of them. A tree is how it is independent of what any of us think about it. Some objects of theoretical reason for me, though, are not independent of everyone's representation of them.⁷⁰ When you act, your action is primarily an object of practical knowledge for you. It depends on your practical representation of it. Yet I can have theoretical knowledge of it. For that matter, when I act, my action is primarily an object of practical knowledge for me. It depends on my practical representation of it. Yet I can have theoretical knowledge of it as well, and it does not depend on *that* representation of it. My own action is not in a theoretical blindspot. Still, theoretical knowledge of actions is of *actions*. To act is to realize a practical representation. Theoretical knowledge of action is thus knowledge of the realization of a practical representation as such. For me to have theoretical knowledge of your action is thus for me to have theoretical knowledge of the object of your practical knowledge *as* the object of your practical knowledge. Anyone can have this kind of

70. If George Berkeley is right, every object depends on God's representation of it. Still, Berkeley need not reject my account of theoretical and practical reason. He need only claim that God's (primary) relationship to an object is not theoretical, which is what the devout should say anyway about a being who is in no way passive or receptive. Anyway, to the extent that I understand the idea of divine perfection, it is incompatible with the idea that God has capacities. A capacity is a potentiality, and a perfect being is in some sense pure actuality. God is thus irrelevant to an account of theoretical and practical reason.

knowledge of action in the sense that the object does not depend on this representation of it. It does not differ in kind when it is knowledge of my action or yours.

In contrast with theoretical reason, though, I have a special relationship to the object of my practical representation. No one else has this relationship to it because practical reason is our kind of will. It is our ability to act, and you cannot make up my mind, though you might try in many ways to get me to make up my mind as you prefer. In other words, an action is essentially the action of its agent because I essentially know my acting in the way that I must know it in order for it to be an action.⁷¹ Again, this claim is about the perfect exercise of practical reason that reveals the nature of that capacity, not about every exercise. Exercises are imperfect to the extent that they deviate from it.

Theoretical and practical reason are capacities that develop over time through training, education, practice, reflection, and everything else that goes into a life. At some point in this development, I come to understand the nature of these capacities. I thereby understand the possibility of internal normative standards applying to their exercises. I then can have theoretical knowledge of practical knowledge—theoretical knowledge of our actions. I can have that theoretical knowledge, though, only because I can act self-consciously, as is characteristic of beings with practical reason. Like my cat, I can act from thought. Unlike him, I can act from thought about action recognized as such. Hence, because theoretical and practical reason are self-conscious capacities that develop over time, I come to be able to act or believe on the basis of grounds that I understand as such. I can act self-

71. I do not mean that an agent must know every aspect of her action in order for it to be an action, nor that an action is defective to the extent that there are any aspects of it that the agent does not know. I exercise many skills in action in ways that might surprise me or escape my notice without thereby acting defectively. I might swerve my bike in order to avoid a snake in the road without consciously registering the snake or having any idea why I swerved, at least not until I investigate the road on my way back home. Yet I exercise these skills in the course of performing an action that I know in the way characteristic of action. It is not like I am just as surprised that I am riding a bike as I am that I am swerving. Such surprise would be the mark of a serious malfunction in our agency.

consciously such that I can answer the questions ‘Why are you doing that?’ or ‘How are you doing that?’ by offering grounds for so acting. I can believe self-consciously such that I can answer the question ‘Why do you believe that?’ or ‘How do you know?’ by offering grounds for so believing. In the good case, my answers constitute genuine explanations of my action or belief that articulate the structure of the exercise itself.

The adverbial version of ‘self-consciously’ is important. I do not mean that alongside my action or belief is a state of awareness or consciousness *that* I am acting or believing and moreover an explanatory account of why I am doing so. I instead mean that we can act or believe in such a way that our self-consciousness is an aspect of the exercise itself. The explanation makes explicit the structure of the exercise. Still, the model of propositional awareness has a place here. For one thing, if I have sufficiently developed my capacity, I am able to reflect on my exercises of practical reason and articulate what I am doing, why I am doing it, and how I am doing it. I am able to come to know *that* I am acting in some way for some reason by articulating *what* I am self-consciously doing. Likewise with respect to exercises of theoretical reason. I need not always reflect, and I need not always be able to reflect in the moment. Doing it might sometimes get in the way of succeeding in the exercise. Still, it is characteristic of this capacity that I can do it. I am in a position to do so in each case that I exercise my capacity perfectly, not about every aspect but about the ones that make it a self-conscious action or belief.

In this way, a third form of reason, reflective reason or reason in its reflective use, comes with the others. The relationship between it and them is complicated. For one thing, part of developing a self-conscious capacity is reflecting on it and its exercises, thereby forming essentially first-personal concepts like ‘reason to act’ and ‘reason to believe’. These concepts are not essentially first-personal in the sense

that they only have application to myself. They are essentially first-personal in the sense that only a being who can self-consciously act and believe can have them. Only a being with theoretical and practical reason can have them. Moreover, anyone who can act and believe self-consciously has everything that they need in order to form these concepts through reflecting on acting and believing self-consciously. In this way, they differ from concepts like GILLS and HORSE, since beings who lack gills and are not horses can have these concepts.

We can form these concepts because we possess theoretical and practical reason. We must form them in order to fully develop these capacities. They mark the explanatory connections I articulate in reflection on an exercise of theoretical or practical reason that constitutes self-knowledge that I am acting in some way or believing something. Hence, the concepts of reasons to act or believe articulate the self-consciousness essential to these exercises. In coming to grasp them through training and reflection, I come to better understand my capacity in a way that allows me to better develop and exercise it. In this way, reflective reason comes with theoretical and practical reason. Exercising it in certain ways is part of developing and exercising them perfectly.

Reflective reason takes as its objects only ourselves and our capacities, including itself and our other self-conscious capacities, and the way that we relate to the world in our exercises of these capacities. Like theoretical reason, reflective reason answers to its object in the sense that a good exercise of it accurately reflects the nature of the capacities in question that are independent of the particular representation of them. Its objects are independent existences in the relevant sense. Unlike theoretical reason, though, the objects of reflective reason are not in general independent of my capacity to represent them in an exercise of reflective reason. Part of developing theoretical and practical reason

includes reflecting on exercises of them, forming concepts through this reflection, and thereby understanding the principles of these capacities themselves. Objects of reflective reason are in this way per se objects of reflective reason in the sense that their perfect state includes reflective development of them and in the sense that their perfect exercises include the use of concepts that we can form only through the exercise of reflective reason. Only a being with reflective reason can have the capacities that are possible objects of reflective reason. They are per se the kinds of objects that we understand through reflection. In this way, they differ from the independent existences that are the objects of theoretical reason. Although we can form concepts through reflection that apply to beings other than ourselves and to beings of different kinds, then, these concepts are essentially first-personal. We can possess them because they can truly apply to us, and they can apply to us because we can guide ourselves by the principles of theoretical and practical reason in exercises of those capacities.

4. A MEANS TO AN END

That is all very abstract and difficult. I shall assume that it is clear and easy, though, and start discussing practical reason in more detail. In this section, I shall explain how the previous account of practical reason establishes an interpretation of the instrumental rule as a standard for exercises of practical reason. This discussion is analytic in the sense that the concept of self-conscious action is enough to establish this standard. In the next section, I will consider whether further standards on action like prudential and moral requirements likewise analytically follow from the concept of self-conscious action.

Practical reason is the self-conscious will. Its perfect exercise is practical knowledge, wherein I realize the object of my practical representation on the basis of that representation. A bit less grandly, I

do what I decide to do, guiding myself in the doing by the representation that I am realizing. I decide to walk to the store in order to get some bread, head down the street toward the store with money and time enough for bread, purchase it, and thereby have bread. From this account of the perfect exercise follows an account of part of principle of our will. The function of practical reason is in part to realize the object of my practical representation. It is to act in a way that realizes the object that I represent because it realizes that object. In order to act in that way, I must act. A bunch of spasms that perfectly resemble such an action may obviate the need to realize a practical representation, but they cannot realize it. After all, the object of a practical representation is an action. I must instead do what I represent myself as doing. The instrumental rule codifies this condition on the successful exercise of our will.

The correct interpretation of the instrumental rule is thereby ‘realize the object of my practical representation’ or ‘take enough means to pull off my end’. This interpretation deviates from familiar interpretations in at least two respects. First, many interpretations treat the principle as about mental housekeeping wherein I must insure consistency among my intentions. It is then about the thinking, not the doing. A brain in a vat could meet it. In contrast, my formulation is the *doing*—about *taking* enough means to the end or *realizing* my practical representation. It is about acting. Second, many interpretations focus on the necessary means to the end, which are sufficient for the end only in unusual circumstances. In contrast, my formulation is about taking enough means to the end, which always includes the necessary means but in most situations goes beyond them. My interpretation, though, accounts for the truth in both of these ideas without divorcing the thinking and the doing that jointly constitute an exercise of practical reason.

Start with the bit about mental housekeeping. The bit about necessary means comes in the

bargain. Because an action is the object of a practical representation, a condition on acting is also a condition on the practical representation of this action. After all, if I am to realize the object of my practical representation and if its object is an action, the representation must include any aspect of the action essential to its being an action. I thereby fail in this representation if I do not represent an action at all or if I represent an action that is impossible for me to do in any of the ways that something is impossible for us. I fail in the thinking if I practically represent standing up and not standing up simultaneously, standing up and sitting down simultaneously, pushing a shadow, salting the number 3, or sailing off of the end of the Earth. In each case, to practically represent the act in question is to fail in this exercise of my will because I cannot realize a representation of an action that is impossible for me. Because I am subject to the instrumental rule in action, only some practical representations are valid or well-formed. That is to say that the function of our will establishes a standard of correctness for my exercises of it that is in part about being able to realize the object of a practical representation. A well-formed practical representation is of something that is not beyond my ability.

Since I have developed my will sufficiently to reflect on it, I know about this aspect of its principle. I act from this knowledge in exercising my will. Most simply, I use it by rejecting options that are beyond my ability. More importantly, though, I use it by determining or specifying generic practical representations. While I often represent something within my ability 'right off', cases wherein I must determine this representation make explicit the structure of a well-formed practical representation. Say that I am to cook dinner tonight for my sweetie in order to express my love for him on his birthday. What to make for dinner? How about a risotto. What do I need to make it? Better look up a recipe online. Need to get the groceries now. Unfortunately, given my habit of always eating out, I do not know

where to get groceries. Back online to find a store and plot my route. Take the route, buy the food, and take it home. Once I get home, though, that damned habit of never cooking gets in the way again. I have no idea how to prepare the ingredients. Fat good they are doing me sitting there on the counter not orchestrating themselves into the right pots and pans! Back to the internet in order to find the recipe again that gave me the ingredients, and then I am off, at least after I call my partner and ask about the whereabouts of our cooking utensils and also how to turn on the range of the stove.

At each stage in this process, I must move from a determinable representation to a more determinate version of it. From cooking dinner for my sweetie in order to express my love for him on his birthday to making a risotto for that dinner, from making a risotto to making this risotto according to this recipe, from planning the meal and finding the store to heading there and buying the ingredients, and so on. At each stage, I continue until I represent something that I have the skill to do in my circumstances. Conscious calculation stops at the point where I can exercise a skill immediately given my surroundings and options.⁷²

This conscious calculation uses my knowledge of causal powers, whether my own or of the objects around me, and of other kinds of facts about the world in general and about the relevant places in the world in particular in order to discover a way to realize my practical representation. This representation, though, is not merely of a way that *someone* can successfully act in this way. It represents a way that is open to *me* to realize my practical representation. A representation of a means that is not a

72. I am characterizing the content or structure of the representation from the first-person perspective. A third-personal description might add epistemic qualifiers to the claims about situations in which I can exercise a skill. For example, 'I continue until I represent something that I have the skill to do in my circumstances' becomes 'I continue until I get to a represent that I *take myself* to have the skill to do in my circumstances *as I take them to be*'. Whatever someone else might say, though, the content of my thought does not include these qualifiers. A characterization of the nature of the will characterizes the agent's thought, not a spectator's thought about the agent's thought.

means for me either because I lack the skill to do it or the circumstances are not right for it is, in the sense relevant to a practical representation, no representation of a means at all. And I know it. It is at best an intermediate representation of the action that I am to realize that I still need to further specify in order for me to be able to do it. A representation of a way to realize my practical representation, then, represents a way that *I* can realize my practical representation *given* my abilities and the world. I need not actually use an ‘I’ in the thought, nor need I to explicitly think about my skills and circumstances. I need only think about the end—making dinner for sweetie—and the means—getting groceries, bringing them home, cooking them up. Yet the representation of a means, even if not explicitly under the concept, betrays my self-knowledge of my skills. To think ‘I need to go to the store now’ is to represent a partial means to my end, where something is a means only if it is the kind of thing that *I* can *do* in order to bring about my end.

‘Means’ is in this sense an essentially practical concept. I need not explicitly use it in order to represent a means to my end, but my representation betrays the structure that comes out explicitly when I say ‘I am going to the store in order to get groceries for the dinner tonight’ or ‘I am cooking dinner for my sweetie tonight in part by going to the store in order to get the groceries for it’. Likewise, the explicit use of the concepts MEANS and END in ‘I am going to the store as a partial means to my end of cooking dinner for my sweetie on his birthday’ articulates the structure of my practical representation. To learn these concepts is to come to understand what I am doing in a practical representation through reflection on practical reason. We can have and use these concepts because we can have means and ends, and we can have means and ends as we do because we can guide ourselves by the principle of practical reason in exercises of it.

In general, a simplified form of this calculation has the following structure.

1. I am ϕ -ing, where ϕ has a means-to-an-end structure but the means are not something that I have the skill to straight off do.
2. Doing A is an at least partial way to ϕ , where A is an act-type that I think partially constitutes the means to my end and that I have the skill to straight off do.
3. I can A in my present environment *in this way*, where 'in this way' picks out a perceptually grasped possibility for doing what I am to do given my skills and abilities.
4. I am A -ing in this way as an at least partial way to ϕ .

I need not actually go through these 'steps' consciously or, really, at all. I often just represent actions straightaway that I can do given my environment and abilities. Maybe I say my name in order to make this very point, where there is no phenomenological gap between the thinking and the doing, let alone between the different steps or stages of representing the action. Yet the structure is still present. I still exercise a skill in a way that at least partially realizes my practical representation given my abilities and circumstances. The steps articulate the form of a practical representation that is present whenever I do anything, even if I do it 'without any thought at all'.

Formulating the instrumental rule in terms of what I must do thus captures everything that the mental housekeeping version captures while also capturing the fact that an exercise of our will that does not include doing anything is thereby defective. Moreover, formulating it in terms of taking enough means to pull off my end rather than taking the necessary means captures the fact that I must achieve my end. The necessary means are generally insufficient for that task, and when they are sufficient it is their sufficiency, not their necessity, that explains why a practical representation that includes only them is well-formed. Again, this condition on the doing brings with it a condition on the thinking because the doing realizes the thinking. A practical representation thus is well-formed only if it represents enough means for my end. The function of our will thus constitutes at least this normative standard for its

exercises. The instrumental rule governs self-conscious agents because of the nature of our will.

This rule also generates further content for the basic normative standard for exercises of practical reason. If each exercise of practical reason is good only if I realize my representation, I must ensure that my practical representations are jointly realizable. So long, at least, as the idea of a capacity brings with it the idea of a bearer of a capacity who can exercise it a manifold of times, the function of our will brings with it this standard independent of any substantive facts about human beings. Hence, the instrumental rule establishes a condition of success for each exercise of practical reason individually and a condition of success for them jointly.

5. CONSTITUTIVE EPISTEMOLOGY

Can the function of our will establish further normative standards on its exercises as such? Constitutivists disagree about whether it can and, if so, which it does. They almost all agree, though, that the only way to show that it can is to analytically derive them from the concept of self-conscious action. I call this view *formal constitutivism* since its central claim is that the form of practical reason is enough to generate the basic normative standards for exercises of the will of any being who can act self-consciously.⁷³ I shall argue in this section, though, that this assumption fits oddly with a constitutivist view of capacities in general. I will consider in the next section arguments for formal constitutivism, or at least criticisms of the alternative *substantive* constitutivism that I develop in this section.

To start, the constitutivist metaphysics does not entail this epistemological condition on normative standards for exercises of capacities. Constitutivism is about the metaphysics of normativity. The assumption about analytic derivability links this metaphysics with an epistemological condition.

73. I take the name from Lavin [*ms.*].

Whence this condition? Nothing in the content of the constitutivism establishes it. Its only epistemological implication is a two way inferential link between knowledge of the nature of a capacity and knowledge of the normative standards that govern its development and exercises. This link says that you can come to know either by correctly inferring it from the other. If you know the nature of a capacity, you know the principle of the capacity in its descriptive role. If you know the normative standard that governs its development and exercises as such, you know that principle in its normative role. Since the same principle plays these two roles, you can infer either from the other.

These inferences fall out of the logic of capacity predication. Perhaps that makes them analytic. However, the assumption about analytic derivability is about something else. After all, although knowing the nature of a capacity is enough for you to know the normative standards that govern the development and exercises of the capacity as such, the metaphysics of capacities says nothing about *how* you come to know the nature of the capacity. In particular, it does not say that the nature of every capacity is analytically derivable from the concept of the activity that differentiates one capacity from another. For all that the metaphysics says, knowledge of the nature of a capacity might be synthetic in the way that knowledge of the nature of water or knowledge of the nature of human digestion is synthetic. Hence, the normative standards that govern the development and exercises of that capacity as such might not be analytically derivable from the concept of the activity that differentiates one capacity from another. To put this point in a way that anticipates the argument in the next paragraph, thinking about this metaphysics of our action within a broader metaphysics of the capacities of the living shows that these normative standards need not follow analytically from the concept of the activity in question.

In fact, thinking about the metaphysics of our action within a general metaphysics of capacities

shows that the normative standards for our action cannot follow analytically from the mere concept of the activity in question. At least, they cannot if they are capacities of the living. Constitutivists tend to focus solely on our action. They tend to talk about rational or self-conscious action, not human action. The species 'self-conscious action' stands to the genus 'action' as the species 'animal nutrition' stands to the genus 'nutrition', and the species 'human action' stands to the genus 'self-conscious action' as the species 'human digestion' stands to the genus 'animal nutrition'. Just as the instrumental rule characterizes self-conscious action, so something like 'take in matter from outside and convert it into matter and energy for use in biological self-maintenance' characterizes animal nutrition. To think that any normative standard that governs exercises of our will as such must analytically follow from the concept of self-conscious action is like thinking that an analysis of the concept of animal nutrition must reveal what every animal should drink and eat and how they should process it.

This task is impossible because of the generality of the genus 'animal nutrition'. 'Animal nutrition' contains all and only what unites the different nutritive processes of the different animals as digestive process. It does not contain what differentiates these processes and requires humans, cows, lions, and mice to eat and drink different things and process them in different ways. These differences come into view only given the substance or specifics of these species of animal digestion. ANIMAL NUTRITION is a *formal* concept. It provides the form of description that the nature of the kind of being in question *substantiates*. Beings with that kind of digestive capacity are thereby subject to a determinate version of the formal rule to take in matter from outside and process it in ways that contribute to their biological self-maintenance.

Given that our will is the capacity of a living being, the explanatory relationship between our will

and the genus of practical reason is the same as the explanatory relationship between our digestion and the genus of animal nutrition. The abstract genus contains all and only what unites different forms of self-conscious wills as self-conscious wills. It does not contain what differentiates our will from other possible self-conscious wills, and it does not tell us or them what to do or how to do it. These differences come into view only given the substance or specifics of the beings in question.⁷⁴ PRACTICAL REASON is a formal concept. It provides the form of description that the nature of the kind of self-conscious being in question substantiates. Beings with that kind of will are then subject to a determinate version of the instrumental rule. I must pursue certain ends in certain ways given my human nature just as I must drink and eat certain things and process them in certain ways given my human nature.

Of course, whereas we are familiar with different kinds of animals with different kinds of digestive systems, we are the only known self-conscious agents. Hence, while we have examples of other specific forms of animal nutrition, we lack examples of other specific forms of self-conscious wills. But so what? The argument does not say that there *are* different forms of self-conscious wills. It does not try to derive the actual existence of kind of living beings from abstract thoughts about the explanatory structure of capacity predication. It only says that different forms of self-conscious wills are possible. In fact, if the lords of grammar allow it, the claim is best put negatively: it is not the case that it is impossible for there to be distinct forms of self-conscious wills. Why? Because a self-conscious will is the will of a kind of living being. The same relationships between nutrition, animal digestion, and human digestion

74. I here part ways with Douglas Lavin about the possibility of a kind of self-conscious agent subject only to the instrumental rule in exercises of its will. On my view, this agent is impossible in the same way that a being subject only to the abstract digestive requirement to take in matter and energy from outside and turn it into matter and energy for use in biological self-maintenance is impossible. Still, Lavin and I agree that different forms of self-conscious wills are possible and on the form of explanation proper to constitutivism. We just disagree about whether that particular one is possible. See (Lavin [*ms.*]) and (Lavin [2004] 428n.3).

would hold in a world in which human beings are the only animals or, if this idea makes sense, a world in which we are the only nutritive beings. They hold because our capacity to digest is the capacity of a living being. The same kind of relationship thereby holds between the will, the self-conscious will, and the human will because our will is the capacity of a living being. The normative standards that govern our action thereby do not all follow from analysis of the concepts of self-conscious action any more than the normative standards that govern our digestion follow from analysis of the concept of animal digestion. The epistemological condition about analytic derivability is wrong. Formal constitutivism is a mistake, and substantive constitutivism is the way to go.

6. KINDS OF WILLS

Substantive constitutivism implies the possibility of alternate forms of self-conscious wills subject to distinct normative standards. All forms of self-conscious wills embody the instrumental rule just as all forms of animal digestive systems embody the principle to take in matter and energy and convert into matter and energy for use in biological self-maintenance. Yet the further substantive norms that these different forms of self-conscious wills embody might differ from each other in the same way that different animals need to eat different stuff and process it differently given their digestive systems. Moreover, even if two kinds of self-conscious wills have the same substantive normative standards, the grounds of these standards differ just as the grounds of the standards for the digestive system for two kinds of animals differ even if they happen to need to eat the same stuff and process it in the same way. They are the same only in content, not in source.

As such, these facts just spell out consequences of substantive constitutivism. Yet I suspect that

formal constitutivists are likely to object to substantive constitutivism on the basis of these consequences in a number of ways. In the next few sections, I shall consider arguments that try to leverage aspects of the self-consciousness of our will against substantive constitutivism. Responding to these arguments lets me clarify the way that the self-consciousness of our will fits into a general substantive constitutivism about capacities. Before discussing these objections, though, I shall in this section offer a model of a non-moral self-conscious will and contrast it with a moral self-conscious will. To be clear, though, representing distinct forms of self-conscious wills is not necessary for the success of substantive constitutivism. Substantive constitutivism entails only that different forms of self-conscious wills are possible. It does not require that we can form a definite positive conception of them.⁷⁵ Still, having these alternatives in view helps me to more clearly present and respond to the objections. For the sake of clarity and cheekiness, I shall use some thoughts from the first division of Immanuel Kant's *Groundwork of the Metaphysics of Morals* in order to characterize these different kinds of self-conscious wills.

Start with a moral agent, by which I mean an agent subject to moral requirements in action as such. The principle of the will of such an agent determines the instrumental rule in such a way that she acts well only if she acts in ways that are morally required or morally permissible. For the sake of argument, say that what determines the instrumental rule is what I call the rule of duty.⁷⁶ A being with such a principle of the will can act from categorical thought about action independent of thought about her antecedent desires. She can act for the sake of justice or friendship, not as the end of her action but as

75. As I explain in §8, Immanuel Kant likewise allows for the possibility of distinct forms of self-conscious understandings because of the nature of the explanatory structure of his view of theoretical cognition without thereby needing or being able to offer an account of the shape of these other forms. See in particular (Kant [1781/7] B150-1).

76. I use 'rule of duty' and 'rule of self-love' instead of 'principle of duty' and 'principle of self-love' for the same reasons that I use 'instrumental rule' rather than 'instrumental principle'.

the principle of choice of an act-end pair that constitutes an action.

This account of the principle of the will of a moral agent is incomplete as an account of our kind of agency. Although it divides possible actions into the required, permissible, and forbidden, it does not rank permissible actions. Unless the principle of the will of a moral agent has more content, then, choices between permissible actions are arbitrary. I do not know whether such a being is possible. I know, though, that human beings are moral beings, and I know that we can act better or worse when acting permissibly. There is a difference between what I must, may, and must not do and what I should or should not do. Although I should do everything that I must do, not everything that I should do is something that I must do. Although I should not do everything that I must not do, not everything I should not do is something that I must not do. Hence, the principle of my will must include more than the rule of duty, something else that ranks permissible actions.

Following Kant, one way to establish such a ranking is to formulate a rule of self-love that is subordinate to the rule of duty that says that an agent is to choose between permissible actions according to which action most satisfies her desires. The principle of the will of a moral agent, then, determines the instrumental rule through the rule of duty and the rule of self-love, with the rule of duty having lexical priority over the rule of self-love. This principle establishes the basic normative standard for exercises of the will of a moral agent. A moral agent who sufficiently develops her will can appreciate this principle and guide herself by it in exercises of that capacity. She acts well if she acts as required or if she acts permissibly and in a way that is least as good as any other permissible action with respect to the rule of self-love. That is, she acts well if she acts as she should act. She acts defectively if she acts in a way that is forbidden or if she acts permissibly but in a way that is less good than another permissible action

with respect to the rule of self-love.⁷⁷ That is, she acts defectively if she acts as she should not act.

Turn now to an egoistic agent. The principle of the will of such an agent, on my interpretation, deletes the rule of duty from this account of the principle of the will of a moral agent. The principle of the will of such an agent thus determines the instrumental rule with the rule of self-love that says to maximize the satisfaction of her desires. As with the moral agent, this rule ranks actions as better or worse depending on how they compare to each other with respect to desire satisfaction. Unlike the moral agent, though, the ideas of obligation, permission, and prohibition do not make sense here. An egoistic agent acts well, then, if she acts in a way that is at least as good as any other action with respect to the rule of self-love. She acts well if she acts as she should act. She acts defectively if she acts in a way that is less good than another action with respect to the rule of self-love. She acts defectively if she acts as she should not act.

7. THE PRINCIPLE OF THE WILL AS OBJECT OF THE WILL

Moral and egoistic wills are incompatible in the sense that no being can be both a moral and egoistic agent. No being can have two wills. Most philosophers, though, treat them as competing accounts of self-conscious agency. On the substantive constitutivist view, this idea is on a par with thinking that human, bovine, and feline digestion are competing accounts of animal nutrition. Philosophers are likely to think that this analogy is mistaken because of the uniqueness of self-consciousness. In the next few sections, I shall consider a number of objections along these lines. Although I take many of these objections from contemporary Kantian views on agency, I cannot do any of them justice. My goal is to

⁷⁷ I am ignoring instances where an agent fails to exercise her capacity to act rather than fails in an exercise of her capacity to act. Nothing important for this discussion turns on this distinction.

use them to get a better sense of the substantive constitutivist view. Let me also be clear about the dialectical situation. The disagreement here is between substantive constitutivists and formal constitutivists, not between substantive constitutivists and Kantians. Standard non-Kantian views of agency in general and standard non-Kantian constitutivists views in particular agree with Kantians that there is only one kind of self-conscious agency. They just disagree about its nature. I consider these objections because of my deep sympathy with the Kantians on many issues that the objections raise and because the Kantian version of formal constitutivist is in my opinion the most plausible, not because the Kantian is the sole enemy of the substantive constitutivist.

Formal constitutivists might first object that an egoistic agent is heteronomous. If a self-conscious will is an autonomous will, an egoistic self-conscious agent is thereby impossible. While the notion of autonomy is a tricky thing to pin down, I shall grant the claim that a self-conscious will is an autonomous will. I instead mean to ask whether a moral agent is autonomous in a way that an egoistic agent is not.

The problem might seem obvious. After all, desires determine the particular standards of correctness for particular exercises of the will of an egoistic agent. Is such determination by something outside of the will not the paradigm of heteronomy? No. The desires of a self-conscious egoistic agent are inputs to a proper exercise of their will because the principle of their will grants them this relevance. They do not establish the basic normative standard for exercises of the will. The principle instead sets the standard. Nor do desires determine the will. Instead, an egoistic agent who sufficiently develops her will can appreciate this principle and guide herself by it in exercises of her will. The normative significance of desires as proper inputs to decision *derives* from the principle of a will. A self-conscious

egoistic agent makes decisions *on the basis* of her desires and act from those decisions. If autonomy is self-determination and if to act autonomously is to guide yourself in thinking and doing by the principle of your will, this role for desire does not rule out an autonomous egoistic agent. After all, the moral agent has proper inputs to decision as well. They determine the particular standards of correctness for particular exercises of his will. Both agents must take into account contexts, circumstances, and other aspects of the world, themselves, and others in order to exercise their will well. It is not as if the moral agent acts in a different world or is in control of more of the world than the egoistic agent. The only differences are which aspects of the world are relevant to decision and the roles that they play. Why should one set of inputs as determined by the principle of the will of the egoistic agent compromise autonomy when another set as determined by the principle of the will of a moral agent does not?

For another thing, the rule that grounds the normative significance of desires as proper inputs to decision for the egoistic agent grounds the same normative significance of desires for the moral agent with respect to decisions between permissible actions. After all, the principle of the will of an egoistic agent is in effect the principle of the will of the moral agent minus the rule of duty. Both agents must guide themselves by the rule of self-love in action in order to act well, the egoistic agent always and the moral agent whenever acting permissibly but not as required. Yet moral agents can act autonomously when acting permissibly. Why, then, would following the rule of self-love compromise the autonomy of the egoistic agent but not the moral agent? If the moral agent can act autonomously even though his desires are proper inputs to decision because of the principle of his will in a limited range of cases, why cannot the egoistic agent in an unlimited range? After all, the rule of self-love is part of the principle of both of their wills.

If the moral agent is autonomous but the egoistic agent is heteronomous, the source of the problem must be something else. My sense is that the worry is difficult to state within the substantive constitutivist framework because the worry in fact takes issue with the way that I treat the principle of the will of self-conscious agents as *given*. On the view behind the objection, the crucial difference between a self-conscious will and the wills of the other known animals is that the self-conscious will is the ultimate practical authority even over its own principle. On this view, I do not *discover* the principle of my will that in fact has authority for my action as such, as the substantive constitutivist claims. I instead *will* this principle, and it thereby has authority for my exercises of my will. As we might put it, I ask the practical question ‘What principle is to govern my will?’, not the reflective question ‘What principle does govern my will?’ I answer that practical question in an exercise of my will, and this act of will establishes the authority of that principle and the autonomy of action from it. In contrast, substantive constitutivism treats the principle of my will of an agent as something that I discover through the reflective use of reason, not something that I establishes through its practical use.

I must admit that I do not understand this view well enough to argue against it on its own terms. Still, let me offer three worries about it. First, establishing the principle of my will in an exercise of practical reason is meant to solve a problem with that principle being given to me. I will question in a moment what it means to establish this principle in an exercise of the will. Even before getting to that issue, though, I do not understand the problem that this view means to solve. Perhaps the worry is that the possibility of different kinds of self-conscious wills implies possessing one kind of self-conscious will rather than another is arbitrary in a way that undermines the authority of the standard that it generates. But in what sense arbitrary? True enough, if I were a different kind of being who by nature

has a different kind of will, then I would be subject to different normative standards. So what? If I were a different kind of animal with a different kind of digestive system, I would need to eat different things and process them in different ways. I am not thereby going to start eating only grass and regurgitating it, chewing it a bit more, and swallowing again because that is how cows do it. Nor am I going to think that my not needing to take up that diet and process things in that way is arbitrary in a way that undermines the standing of the standard for human digestion with respect to me. To do so would be the mark of a deep psychological problem, not evidence that normative standards that my digestive capacity establishes for its exercises lack authority for my digestion as such. The problem would be with me, not with the standards. The solution is therapy, not a revision of a philosophical view about the nature of normativity. That is the way to deal with alienation from human nature, whether it is alienation from our digestive capacities or from our will.

Second, the view that I establish the principle of my will in an exercise of my will is problematically metaphorical. For one thing, I never actually ask myself this question about what is to govern my will. For another thing, for long periods of my life, I am unable to ask myself this question because I have not sufficiently developed my will. Practical reason does not come fully formed in us. It takes work to be able to exercise it well, and it takes work to be able to reflect on it in the way that asking and answering this question requires. Since formal constitutivists do not think that we are not subject to the principle of our will until we actually, if ever, address this question, the metaphor is ill-chosen. Yet formal constitutivists who criticize substantive constitutivism along these lines need this kind of metaphor in order to distinguish their view from the view that the principle of the will is *given* in the sense that they take to incompatible with autonomy.

Can they spell out the metaphor in a way that is compatible with the facts about our will on which I base my explanation and defense of substantive constitutivism without saying that we discover rather than establish the principle of our will? I doubt it. To take just one example, they might try to explain the metaphor in terms of the *form* of an exercise of the will that restricts the *matter* of such an exercise in certain ways in the way that the instrumental rule says that the object of the will must have means-end or part-whole structure. I endorse this way of thinking about the principle of the will, but how does it fit with the metaphor? After all, although every exercise of the will *bears* this form, this form is no more willed than the form of human digestion that all of my exercises of that capacity bear is digested. Moreover, even if the will could take this form as its object and will it, *that* exercise of the will bears that form, not as matter but as form. Its role as form and thus as unwilled restricts the content of this exercise of the will as it does any other. Its authority is thereby not established by this exercise or any other. It is instead presupposed by the exercise. Hence, unless formal constitutivists spell out the metaphor in another way, it is misleading and is no support for formal constitutivism.

Third, whatever comes of that worry, formal constitutivists must think that there is a correct and incorrect way to answer this question. The basic structure of the will is not a liberty of indifference that is only subject to a standard of correctness after it arbitrarily wills a principle for itself. Otherwise, there are a manifold of different forms of self-conscious wills, one for each possible arbitrary decision. What establishes the standard of correctness for this exercise of the will? It cannot be the principle of the will. Its authority depends on this exercise of the will. It cannot be something other than the principle of the will or this exercise of the will. Otherwise, the will is not autonomous in this act but instead takes instruction from elsewhere. It thereby is not autonomous in any of its exercises. Heteronomy corrupts.

Heteronomy of the principle of the will corrupts absolutely. The only option is that this exercise of the will establishes the normative standard to which it is subject as such. However, different answers to the question then establish different normative standards on themselves.

Perhaps formal constitutivists think that this result is not a problem because only one answer meets the standard that it establishes for itself. In this way, the question has a single correct answer despite the fact that any answer to the question sets up the standard of correctness for itself. Absent a very tricky argument, though, this response is unpromising. Presumably, a decision meets the standard that it establishes on itself only if acting from that principle in some sense succeeds *on its own terms*. After all, the principle of a moral will is not ruled out because hurricanes can disrupt and defeat obligatory actions. I do not know about how formal constitutivists means to establish this condition that the principle of the will succeed on its own terms. Perhaps it is an expression of the instrumental rule, which follows from the concept of self-conscious action. Yet that rule is not willed by anyone, at least without presupposing it. Let that issue pass, though, and focus on a decision succeeding on its own terms. A commitment to chaos is as consistent with itself as a commitment to order, a commitment to malevolence as approving of its realization and likely effects as a commitment to benevolence. Whence the failure of one and the success of the other against the standards that they establish for themselves? This question is harder to answer than it at first appears. After all, anything a formal constitutivist invokes in response to it is relevant for an agent only if the agent wills *it* in response to the basic practical question 'What principle is to govern my will?' Otherwise, formal constitutivists have no criticism of the substantive view that treats the principle of the will of an agent as something that she discovers through the reflective use of reason, not something that she establishes through its practical use. Yet anything

that they invoke that might seem to explain the difference is subject to the same problem. Somehow it and only it must survive its own appraisal—that is, somehow it must survive its own appraisal and everything else must fail theirs.

Obviously, I am merely offering a schematic way to respond to this kind of objection. I cannot go through every possible argument and show how to apply this form of response. I cannot even consider the Kantian view that the principle of the will of a moral agent embodies ‘universality’ or ‘the form of lawfulness’ in a way that distinguishes it as autonomous and rules out other kinds of self-conscious wills. Such consideration requires a separate paper solely devoted to that issue. I shall instead move to other issues that formal constitutivists are likely to have with substantive constitutivism. Still, I think that the discussion to follow touches on many of the issues that considering the familiar view about universality would, though perhaps in a different form with different emphases.

8. THE UNITY OF WILL

In this section, I shall consider an objection that says that substantive constitutivism is incompatible with a certain publicity or shareability condition. I will focus on a version of this criticism that I draw from Stephen Engstrom’s *The Form of Practical Knowledge*.⁷⁸ The objection is that the possibility of distinct forms of self-conscious wills is incompatible with a certain condition on the possibility of knowledge. Since practical reason is the self-conscious will and its perfect exercise is practical knowledge, substantive constitutivism would then be incompatible with the possibility of a self-conscious will.

In order to understand the condition on the possibility of knowledge, start with theoretical

78. Although I do not discuss them, I think that Christine Korsgaard’s claims about the publicity of reasons and Andrews Reath’s claims about the public availability of any valid instance of practical reasoning embody the same basic thought. See Korsgaard [1993; 1996, part 4] and Reath [1994; 2006_b].

knowledge as the perfect exercise of theoretical reason. Say that I know that there is a tree in the courtyard. Theoretical knowledge is of an independent existence. Given a sufficiently developed capacity of theoretical reason, I know that this tree is independent of my representation of it. I thereby know that other beings with appropriate capacities to know can come to share this knowledge of the tree in the same way that I came to it. They can stand in the same kind of relationship to the tree. Moreover, I can transfer this knowledge to other beings with theoretical reason through testimony even if they do not see it. In effect, the same kind of knowledge of the tree is shareable between different subjects. A unity of theoretical knowers is possible with respect to any object of theoretical knowledge.

The objection turns on constructing a similar claim about practical knowledge. To put the point in somewhat grand terms, just as theoretical knowledge is possible only if there can be a unity of the understandings of different subjects, so practical knowledge might seem to be possible only if there can be a unity of the wills of different subjects. The claim that substantive constitutivism cannot account for this unity of wills can take two forms. One criticism says that a unity of wills is impossible because of the possibility of distinct forms of self-conscious wills. Another says that only the moral form of a self-conscious will can unify the wills of its bearers. I shall take up these criticisms in such a way that the second criticism responds to the failure of the first.

The first criticism says that self-conscious agents with different kinds of wills cannot unify their wills in the sense necessary for practical knowledge. Since practical knowledge is the perfect exercise of the self-conscious will, there cannot be different forms of self-conscious wills. Here is a way into the criticism. The principles of the wills of moral and egoistic agents result in different classes of right and good actions. With respect to at least some actions in at least some circumstances, a moral agent and an

egoistic agent exercising their wills perfectly do not agree on the normative status of an action. They thereby cannot unify their wills with respect to it. Moreover, even if they agree with respect to the status of an action in a circumstance, the ground of the goodness or rightness of the action will be different. Surface unity about the status of the action masks deeper disunity about the source of that status. If unity of wills requires agreement in verdict and agreement in the grounds for that verdict, moral and egoistic agents with different kinds of wills thus cannot unify their wills. Yet if every object of practical knowledge requires the possibility of this unity of wills, these agents cannot have practical knowledge. If they cannot have practical knowledge, they cannot have self-conscious wills.

Although in a moment I shall challenge the claim that a unity of wills is necessary for practical knowledge, let it go for a moment. I here wish to point that this criticism is ambiguous. Start with the claim that a unity of knowers is possible with respect to any object of theoretical knowledge. This claim might mean that any possible object of theoretical knowledge for some subject is a possible object of theoretical knowledge for at least one other subject. It might instead mean that any possible object of theoretical knowledge for some subject is a possible object of theoretical knowledge for every subject. The claim about the unity of wills is likewise ambiguous. It might mean that any possible object of practical knowledge for some subject is a possible object of practical knowledge for at least one other subject. It might instead mean that any possible object of practical knowledge for some subject is a possible object of practical knowledge for every subject.

The mere possibility of distinct forms of self-conscious wills is only incompatible with the stronger interpretation, but the stronger interpretation is implausible even with respect to theoretical knowledge. To understand why, consider Kant's claim that the

pure concepts of the understanding are related through the mere understanding to objects of intuition in general, without it being determined whether this intuition is our own or some other but still sensible one, but they are on this account mere **forms of thought**, through which no determinate object is yet cognized. The synthesis or combination of the manifold in them was related merely to the unity of apperception, and was thereby the ground of the possibility of cognition *a priori* insofar as it rests on the understanding, and was therefore not only transcendental but also merely purely intellectual. But since in us a certain form of sensible intuition *a priori* is fundamental, which rests on the receptivity of the capacity for representation (sensibility), the understanding, as spontaneity, can determine the manifold of given representations in accord with the synthetic unity of apperception, and thus think *a priori* synthetic unity of the apperception of the manifold of **sensible intuition**, as the condition under which all objects of our (human) intuition must necessarily stand, through which then the categories, as mere forms of thought, acquire objective reality, i.e. application to the objects that can be given to us in intuition, but only as appearances; for these alone are we capable of intuition *a priori*. (Kant [1781/7] B150; underlining added)

A full analysis of this passage is not my game here. The important part is the connection between the two underlined bits. The categories are the concepts that must apply to any possible object of theoretical knowledge for some subject. For any particular subject, the categories only apply to objects as appearances, and the form of sensibility constitutes the form of objects as appearances. Hence, appearances are the only possible objects of theoretical knowledge. However, at the start of this passage, Kant explicitly allows for the possibility of forms of sensibility that are distinct from the spatiotemporal form that characterizes human receptivity. Possible objects of theoretical knowledge for human beings are possible objects of theoretical knowledge for other human beings. However, they are not possible objects of theoretical knowledge for some other possible self-conscious subjects.⁷⁹ And this Kantian view is in fact an abstract and basic version of the familiar view that different kinds of self-conscious agents with different capacities to know might not be able to share knowledge. I and other humans can know

79. Kant's distinction between real possibility and merely logical possibility is relevant to the proper interpretation of this claim. I ignore it here, though, because substantive constitutivism only commits itself to the claim that we cannot rule out the possibility of alternate forms of self-conscious wills. The use of the egoistic form of a will is for expository purposes, and the basic point about the explanatory structure of the view holds even if the egoistic agent is in the end not a real possibility.

that the chair in front of me is blue. Self-conscious beings who by nature lack color perception cannot. Still, I can know the colors of things. Hence, although any possible object of theoretical knowledge for a self-conscious subject is a possible object of theoretical knowledge for at least one other such subject, they need not be possible objects of theoretical knowledge for every self-conscious subject.

The possibility of distinct forms of self-conscious wills likewise implies that possible objects of practical knowledge of a self-conscious subject are not possible objects of practical knowledge for all such subjects. Yet it does not by itself show that possible objects of practical knowledge for a self-conscious subject are not possible objects of practical knowledge for any other self-conscious subject. Formal constitutivists need a different criticism in order to show that substantive constitutivists violate that interpretation of the condition on the possibility of knowledge.

Formal constitutivists should here appeal to the second way to understand the original worry based on the condition on knowledge. This criticism says that although the principle of the will of a moral agent allows for the possibility of a unity of wills, other principles do not. Hence, only the principle of the will of a moral agent allows for the possibility of practical knowledge. Since the perfect exercise of a self-conscious will is practical knowledge, self-conscious agents are moral agents.

For this argument to succeed, formal constitutivists must show that the principle of the will of a moral agent allows for the unity of wills and that no other principle does. Since proving a negative existential claim is hard, this task might seem insurmountable. Substantive constitutivists might seem to be able to take refuge in that fact. However, I shall argue against the criticism more directly by showing that the idea of a unity of wills is not a condition on practical knowledge. The condition on knowledge makes sense with respect to theoretical reason because of aspects of the relationship between a

representation and its object that distinguish theoretical reason from practical reason. It does not make sense with respect to practical reason because of the *practicality* of a practical representation. It is a condition on theoretical knowledge, not practical knowledge. So I shall argue.

Let me first explain why someone might think that moral agents can unify their wills in a way that egoistic agents cannot. The principle of the will of a moral agent first filters actions into the required, permissible, and forbidden. Only then can the idiosyncratic desires of the agent in question play a role in determining which among the permissible actions is best for the agent. Whereas the idiosyncratic desires of the agent play a role in this second stage of thought, they do not have a role in the first stage. Rather, the rule of duty applies to the same situation in the same way regardless of the idiosyncracies of the moral agent in question. As we might put this point, the rule of duty applies to the agent only under a representation of her as *a* moral agent, not under her representation of her as *this* moral agent. The rule of self-love applies to her as *this* moral agent. The rule of duty is in this way impartial while the rule of self-love is not.

This impartiality might seem to allow for a unity of wills of moral agents. Moral agents who exercise their wills well will appreciate the same actions as required, permitted, or prohibited on the same grounds. In contrast, the principle of the will of an egoistic agent involves the idiosyncratic desires of the agent from the start because the rule of self-love applies to her as *this* egoistic agent. This partiality might seem to preclude a unity of wills of egoistic agents in the sense that agents who exercise their wills well will not appreciate the same actions as having the same normative status on the same grounds.

This explanation is no doubt incomplete, but it is enough for me to show why a unity of wills is not a condition on practical knowledge. To start, recall that the principle of the will of an egoistic agent is

the principle of the will of a moral agent minus the rule of duty. Any problem with the principle of the will of the egoistic agent in general is then also a problem for the principle of the will of a moral agent with respect to permissible actions. If a unity of wills is possible only if different agents agree with the normative status of the action in question on the same grounds and if the rule of self-love is incompatible with this unity, it is incompatible with this unity when it determines the relative normative status of permissible actions for the moral agent. At best, a moral agent can think about the permissible action of another moral agent and understand the normative status of the action given the desires of the other moral agent. This understanding is not an exercise of the will, though, because the principle of the will tells the moral agent to decide between permissible on the basis of her desires, not the desires of the other agent. It is theoretical knowledge of practical knowledge, not practical knowledge, and one egoistic can have this same kind of knowledge of another egoistic agent. In fact, a moral agent can have theoretical knowledge of the practical knowledge of an egoistic agent. He can have theoretical knowledge that her action meets the principle of her will even if his will has a different principle in the same way that he can understand that a cow is digesting well even if his digestion has a different principle.⁸⁰

Moreover, the possibility of this kind of theoretical knowledge of practical knowledge of the action of another agent is the only plausible condition on the possibility of practical knowledge. Practical knowledge is productive knowledge. A practical representation is of a dependent existence, of an object

80. Although this point does not matter for my argument here, I do not think that an egoistic agent can have theoretical knowledge of the practical knowledge of a moral agent. I do not think that she can have the concepts of obligation, permission, and prohibition because of the essential first-personal nature of concepts of practical normativity. An egoistic agent can operate only with concepts of the good, not with concepts of the right that mark the characteristic kind of goodness possible for a moral agent. I say more about this stuff in the next section.

that is per se the object of that representation and realizes it. You cannot stand in this relationship to my action. My action cannot realize your practical representation since it is the object of my representation, not yours. To think otherwise erases the distinction between the agent of an action and spectators of that action. It erases the distinction between the self-consciousness of action and the way that you know about my action by observing it or otherwise coming to know about it. The unity of knowers is plausible in the case of theoretical knowledge because theoretical knowledge is of a distinct existence. My theoretical knowledge of something is of an object that exists independently of that representation of the object. It is thereby of something that another can represent in the same way. My practical knowledge is not. It is of dependent existence. It is thereby of something that others cannot represent in the same way.

Far from being a condition on the possibility of practical knowledge, the unity of wills seems plausible only if you ignore the aspects of practical knowledge that distinguish it as practical. It seems plausible only if you think that the basic relationship in which I stand to my action is of a kind with the relationship in which you can stand to my action. Far from a mark of practical knowledge, this relationship is essentially alienated. To insist on the unity of wills is thereby to miss the self-consciousness essential to practical reason. It is incompatible with the possibility of such a capacity.

The moral agent and the egoistic agent differ in an important respect that might seem to bear on this issue but in fact does not. Moral agents can act together. They can act as joint agents by unifying their wills. Egoistic agents cannot. The grounds of their decisions are distinct given the indexical in the principle of their wills. Hence, shared practical knowledge is possible for moral agents but not egoistic agents. Yet this fact just reflects the basic difference between these kinds of agents. As we might put the point, whereas moral agents are cooperative agents who compete only for the sake of cooperation,

egoistic agents are solitary agents who cooperate or compete only for the sake of solitary gains.

Yet this isolation of each egoistic agent from other self-conscious agents does not matter for the possibility of practical knowledge. The possibility of shared practical knowledge cannot be a condition on the possibility of practical knowledge. Otherwise, we cannot succeed in actions that by nature have only one agent. Kant might think that masturbation is immoral and thus not a possible object of practical knowledge, but this view is preposterous. Yet masturbation is not the object of practical knowledge for more than one agent. That would be a different kind of sex act altogether. Others can know about it, as embarrassed teenagers learn. Yet they cannot know it in the way that the agent does. Every egoistic action is in this way like masturbation and unlike sex between consenting moral agents. Yet the possibility of shared practical knowledge for moral agents no more establishes the impossibility of solitary practical knowledge for egoistic agents than the possibility of sexual congress establishes the impossibility of masturbation.

Whereas a possible object of theoretical knowledge for an agent, then, must be a possible object of theoretical knowledge for at least one other agent, a possible object of practical knowledge for an agent need not be a possible object of practical knowledge for at least one other agent. It must be a possible object of theoretical knowledge for any other agents who can know about it. It can be a possible object of practical knowledge for at least one other agent only if the agents in question have the appropriate kinds of will and the object in question is a joint action. This difference between theoretical and practical reason reflects the difference between how practical and theoretical representations relate to their objects. Many beings can stand in the same kind of relationship to an independent existence, whereas only an agent of an action can stand in the same kind of relationship to a dependent existence. A

dependent existence can only depend on the practical representation of it.

Another way to this conclusion is to reflect on a difference between truth and goodness. The truth of something rules out the truth of anything that conflicts with it. Put otherwise, the truth of something entails the falsity of anything that conflicts with it. The goodness of something does not likewise rule out the goodness of anything that conflicts with it. Put otherwise, the goodness of something does not entail the badness of anything that conflicts with it. Think of equally good but incompatible means like distinct routes to the same destination or equally good but incompatible ends like different exhausting but fulfilling careers. While these instances in which goodness of something does not rule the goodness of other things that conflict with it happen within one action or within sets of options for one agent, the same possibility opens up between the actions of different agents. Think about the other animals. Catching the antelope is good for the lion. Escaping the lion is good for the antelope. The conflict does not rule out the possibility of antelopes and lions as distinct kinds of beings with distinct ways to act and be. It only shows that what is good for one of them need not be good for the other.

The possibility of an object of practical knowledge for one agent that is not a possible object of practical knowledge for another is like the possibility of an action being good for the lion without being good for the antelope. Just as the goodness of these respective actions is not shareable between the lion and antelope, so practical knowledge of an action might not be shareable between one agent and another. It might not be shareable because the agents have different principles of their wills in a way that mimics how the lion and antelope are different kinds of animals. That is how it is with respect to the moral and egoistic agents. It also might not be shareable because the agent have the same principle of the

will but it does not allow for the unity of wills in a way that mimics how the goodness of one lion taking over leadership of a pride is bad for the lion who loses the position even though they are the same kind of animal. That is how it is with respect to egoistic agents.

One last way that formal constitutivists might try to establish the need for the possibility of the unity of wills with respect to a practical object is to say that I must be able to will any condition needed in order to ensure as much as possible the success of my action. Given my finitude, one thing that I need is for other agents to not get in my way and muck things up. Moral agents will not get in the way when other moral agents exercise their wills well and perform permissible or required actions. Egoistic agents, though, might muck each other up since what would maximize satisfaction of my desires might not maximize satisfaction of your desires. Perhaps moral rational agents are possible but egoistic ones are not because willing cooperation is implicitly part of every exercise of a self-conscious will. In fact, perhaps such an will for cooperation is part of the instrumental rule. Just as that rule requires me to take sufficient means to my end, perhaps it requires me to will that other conditions likely needed to bring about my end hold.

However, this argument turns on misconstruing the nature of the will. With respect to things that are proper objects of my will and within my control, I must will them in order to ensure as much as possible the success of my action. That is to say little more than that the instrumental rule is part of the principle of my will. I do not *will* gravity or friction, though, even though they are conditions needed in order to ensure the success of almost all of my actions. Try walking somewhere without them. It is a mark of finitude that we act in a world not under our control and on matter that we need in order to act whose properties are not entirely up to us. It is a mark of the separateness of persons that you are not

under my control. To say that I must will gravity, friction, or your cooperation is at best to confuse what I must *hope* to be the case with what I must *will* to be the case. It is at worst to confuse a state of megalomaniacal madness for a condition on rational agency. Gravity, friction, and how you make up your mind are not the kinds of things that I in general can or cannot will despite often being needed in order for me to act at all, at least in the ways characteristic of human beings.

Before I move onto the next objection, I want to explicitly deny that this argument implies that self-conscious agents cannot share practical knowledge if they have different principles of the will. They have different principles of their wills if they are different kinds of animals. These principles might conflict in a way that does not let share them practical knowledge. However, they might not. If the principles of their wills are compatible, they can share practical knowledge. For example, imagine that I promise you that I will ϕ , but we discover that I am in fact a neanderthal, not a human being. The promise is not thereby null and void. Despite our being different kinds of animals, our ability to engage in categorical thought together lets us to act together and share practical knowledge. Hence, humans can stand in moral relationships with some other kinds of self-conscious beings even though they need not be able to stand in them with every kind.

Say that for a self-conscious agent to be a person with respect to another such agent is for them to have reciprocal moral standing with each other. Some self-conscious agents, then, can be persons with respect to the other members of their kind. Some of them can also be persons with respect to some other kinds of self-conscious animals. Yet other self-conscious agents cannot be persons with respect to anyone. We are persons with respect to each other. We might be persons with respect to self-conscious Martians. We might not. It depends on what they are like in much the way that whether an animal of one

kind can mate with one of another depends on what they are like. Tigers can mate with lions but not elephants. Horses can mate with donkeys but not giraffes. Perhaps humans can get on with Martians but not beings from the dark side of the moon. We will not know unless we meet them and try it out.

9. REFLECTION AND THE PRINCIPLE OF THE WILL

The criticism that I discuss in §7 says that one issue with substantive constitutivism is that it treats the principle of my will as something that I discover through the reflective use of reason, not something that I establish through the practical use of reason. The challenge that I discuss in this section shares with that criticism the suspicion that substantive constitutivism cannot show that I can come to understand the principle of my will through the reflective use of reason. However, this challenge does not find fault with the bit about the reflective use of reason. It instead finds fault with the idea that this kind of reflective knowledge fits within the substantive constitutivist framework. It in effect argues that I know the principle of my will through the reflective use of reason but that I can only have reflective knowledge of the aspects of my will that follow from the concept of self-conscious action. Since formal constitutivism says that these aspects exhaust the content of the principle of my will and substantive constitutivism says that it includes more than them, I can know the principle of my will through the reflective use of reason only if formal constitutivism is true.

Start with the way that we come to understand principles of non-self-conscious capacities. I come to understand the principle of my digestive capacity, say, by empirical means. I observe exercises of it in myself and other humans. I separate the good exercises from the bad ones depending on whether they contribute to biological self-maintenance and figure out the properties of the good exercises that

separate them from the bad. I thereby come to understand the nature of this capacity through the successful exercises of it. These exercises are independent of my knowledge of them and of my conception of what makes them good or bad. My knowledge of them is theoretical knowledge through an empirical method. In this way, my knowledge of human digestion is not different in kind from my knowledge of the nature of the bovine or the feline digestion.

However, formal constitutivists says that I cannot come to understand the nature of my will through this empirical method. Because our will is a self-conscious capacity, I cannot observe exercises of it, see if they achieve their goal, and thereby come to understand the nature of the capacity through the successful exercises of it. In order to understand the problem, think about the essentially first-personal nature of the concept REASON TO ACT. I come to possess this concept in developing my will. This development crucially involves training in exercising my will, which includes instruction in the ways I am to act given my elders' sense of the normative standards for our action. In other words, my elders instruct me in the exercise of my will based on their conception of the normative standards for exercises of our will. Part of developing my will is thus coming to a substantive conception of the normative standard for exercises of our will and thus a substantive conception of the normative standards for our action. This substantive conception will not be identical to the substantive conceptions of my elders because part of developing this capacity is developing critical and reflective abilities. Still, it is a substantive conception of the normative standards for exercises of our will.

To that extent, developing my will might not seem too different from developing my digestive capacity. After all, my elders tell me what to eat. Eating is a kind of doing and a part of digestion. Yet our digestion differs from our will because I guide myself by my conception of the normative standards for

exercises of our will in acting. I also use those standards in order to evaluate the actions of other human beings. My understanding of the principle of my will thus plays a role in my exercises of that capacity that my understanding of the principle of my digestive capacity does not play in my exercises of that capacity. What I eat might depend on my sense of what to eat, but the rest of digestion is independent of my sense of how it is to go. In contrast, what I think about what to do, what I do, and what I think about what I and others do is not independent of my sense of what to do. My sense of the principle of my will feeds into the exercise of my will, my observation of them, and my observation of other human exercising their will in a way that the conception of the normative standard for our digestion does not.

Because of these aspects that distinguish our will as a self-conscious capacity, the empirical method by which we understand the nature of our digestive capacity might seem to be inappropriate for understanding the nature of our will. Humans with different senses of the normative standards for our action can agree about whether someone meets the instrumental rule while the one finds virtue, the other vice, in the action. The exercises and observations do not have the necessary independence from our understanding of the nature of our will in order for them to function as tests and correctives to that understanding. The empirical method thereby does not seem to work with respect to our understanding of our will.

On its own, this difference is not a problem for substantive constitutivism. Fitting our will into a general metaphysics of capacities does not require denying any differences between self-conscious and non-self-conscious capacities. It only requires explaining those differences within the general metaphysics. To show that we cannot understand the nature of our will through the empirical method is not to establish that we cannot understand it anymore than to show that you cannot get to Manhattan

on Route 66 is to establish that we cannot get there. It reveals a problem only if we cannot otherwise understand the nature of our will.

The empirical method is the only way to understand the nature of our will within a substantive constitutivist framework only if we cannot learn about substantial aspects of our will through reflection. What is the same thing, it is the only way to understand the nature of our will within this framework only if we can only learn about the formal aspects of our will through reflection. After all, formal constitutivists are not sceptics about the possibility of knowledge of the nature of our will. The question, then, is not whether we can learn about our will through reflection. It is instead about what we can learn in that manner. Formal constitutivists might claim that the answer is obvious. If you can analytically derive something from the concept of something else, you can learn about the first thing merely by reflecting on the second. The scope of reflective knowledge, then, might seem to be identical with the scope of knowledge by this kind of analytic derivation. Yet this view is implausible for two reasons.

First, we come to know *that* we have a self-conscious will through reflection. Such reflective knowledge is not analytic. Whence would we derive it? Nor is it empirical. It is not as if I observe a series of exercises of my will and thereby posit that I have a self-conscious will in the way that I observe a series of exercises of my digestive capacity and thereby posit that I am omnivorous. If I could learn about my will in that way, it would not be self-conscious. Instead training, instruction, and other things bring me to a point in the development of my will where I can reflect on it and thereby realize that I can exercise it self-consciously. Hence, any scepticism that we can know the substantive aspects of our will that distinguish it from other possible self-conscious wills cannot derive from a general view that reflective knowledge is analytic knowledge.

Second, if every aspect of the principle of our will is analytically derivable from the mere concept of self-conscious action, the skills whose exercise constitutes the exercise of our will are alien to the will. After all, we cannot simply exercise our capacity to act. We instead exercise our capacity to act by, say, walking or jumping. The only way to develop and exercise the determinable capacity is to develop and exercise determinates of it. Yet none of these more determinate activities, let alone the standards that govern exercises of the determinate capacity to engage in them, follow analytically from the concept of self-conscious action. Any self-conscious will is in part a capacity for self-movement, but the kinds of movement will depend on the kind of being in question. After all, self-movement is like self-conscious action in being a determinable capacity that I can develop and exercise only by developing and exercising determinates of it. Yet if every aspect of our will must follow analytically from the concept of self-conscious action, these determinate capacities must be external to the will. They must figure in practical thought as part of the circumstances in which I exercise my will, not constituent elements of my will as I have developed it that I exercise. I then am alienated from my skills and my body, relating to them as I relate to the ground on which I walk and the water in which I swim. To not be alienated from them in action, they must be part of the determinate way that I develop my will. My basic capacities to act in some way that come with being a human being must be part of the nature of my will.⁸¹ Yet these

81. My non-basic capacities to act in some way must be part of my developed state of my will, but they need not be part of the nature of my will in the same way that my capacity to speak language is basic and part of what it is to be a human being but my capacity to speak English is not. I by nature have a capacity to act and the more determinate capacity to walk in the way that I by nature have a capacity to speak language. I do not by nature have the capacity to play video games or a capacity to speak English even though I can train into those activities given my basic capacities. They can be parts of my developed will and intelligence even though they are not part of the nature of the will. I do not have an account of which determinates come with our nature and which do not. The important question is whether not developing the determinate constitutes a deprivation or defect. A rough test is to think about whether the activity in question is something that any of us must be able to do in order to live a human life well. This test is not foolproof. It relies on a conception of our kind of life as present in the many ways we live that distinguishes the essential and non-essential aspects of those ways of life. I do not have a procedure to recommend for answering these questions other than 'think hard and carefully'.

capacities cannot be analytically derived from the concept of self-conscious action. Otherwise, it would be analytic that all self-conscious beings walk, jump, speak, and the rest of it as human beings do. Since human beings are a particular species that evolved on a particular planet at a particular point in time, such a dramatic anthropocentrism would be a rather dull end to the search for intelligent life in the rest of the universe. Such hubris is I take it no longer possible. At least, it is out of fashion.

Reflective knowledge, then, need not be analytic knowledge. Moreover, the way that I know that I can walk, talk, jump, and all that jazz self-consciously provides a model for how I can know the other aspects of the principle of my will. I cannot know this stuff without training, without developing my will to the point where I can reflect on it. Yet this knowledge is not empirical. It is not as if once I get to that point I am able to observe the exercise of my will and discover things about it in the way that if you cut me open, drug me up, and set up some mirrors I am in a position to observe the workings of my digestive system and discover stuff about it. The knowledge is not *external* to the skill in the way that the knowledge of digestion is external to it. The knowledge is built into the skill when it develops as it by nature develops. To put the point in the opposite way, the kind of knowledge of walking built into the skill of a self-conscious agent is internal to the skill. It is a skill that I can exercise self-consciously—a skill *per se* to be exercised self-consciously—and I understand it as such in possessing it in this developed state.

Knowledge of skills like walking might seem a far cry from knowledge of the aspects of the principle of the will that are the primary focus of ethical theorizing. But maybe not. After all, one thing I can know about my will through reflection is that I act from categorical thought rather than only through merely calculative thought of the kind characteristic of the egoistic agent. I know that I can fulfill a promise or pay a debt regardless of whether they satisfy any of my desires. I know that your good can

be a ground for my action independent of it having any connection to my good. I have more and less determinate concepts of obligation, permission, and prohibition. I herein have self-knowledge through reflection on my self-conscious capacity that distinguishes me from the egoistic agent who cannot act in these ways and cannot possess these concepts. This self-knowledge is like knowing that I can have color knowledge that distinguishes me from possible knowers without visual systems.

I likewise can reflect on the conditions for the possibility of exercising my will in various ways. With respect to forms of self-movement for humans, I recognize the need for matter on which I act. I need a suitably strong surface in order to walk. I need a suitably liquid environment in order to swim. No walking or swimming in a zero-gravity vacuum environment. Likewise, I can recognize that I can act on a promise only if the other accepts and I can act with another in general only if shared practical knowledge is possible for us. Understanding these capacities and the conditions of possibility for their successful exercise lets me understand the nature of my will because their developed state embodies a kind of understanding of them and thus a kind of self-understanding. It is an imperfection of the capacity to walk for a human being to be able to walk and yet not understand the skill in a way that lets her have self-knowledge of this ability as the developed state of a self-conscious capacity. Less grandly, it is an imperfection of this capacity to not know that you can exercise it self-consciously.

Such self-knowledge is substantial reflective knowledge. It is enough for us to know that we are not egoistic agents since such agents cannot act from categorical thought. Still, is it enough for us to know that we are moral agents? After all, acting against your good can be as categorical as acting for your good. What shows that one of them is a perfection, the other an imperfection, of an exercise of my will? After all, the metaphysics of capacities says that the capacity brings with it two subclasses of possibility,

what I can do because I have the capacity and what can happen because I have it. What I can do marks perfections, what can happen imperfections, of exercises. If our capacity to act is in part a potentiality to act from categorical thought, it distinguishes the good instances of thus acting from the bad ones. Yet I can know that acting for your good is in one subclass of what is possible and acting against it the other without knowing to which subclass each belongs. Is the 'can' in 'I can act for your good' the 'can' of what I can do given my capacity to act or the 'can' of what can happen given that capacity?

Try it this way. Whereas egoistic agents are solitary agents and moral agents are cooperative agents, perhaps immoral agents are competitive agents. Egoistic agents are out for themselves, moral agents look out for each other, and immoral agents are set against each other. If the 'can' in 'I can act for your good' is the 'can' of what I can do given my capacity to act, I am a moral agent. If it is the 'can' of what can happen given that capacity, I am an immoral agent. Which is it? How can I know?

The possibility of ignorance is of course dull. Any account of the knowledge of a finite kind of agent must allow for ignorant members of the kind with respect to most objects of knowledge, whether reflective, theoretical, or practical. The question, though, is about the possibility of knowledge. Yet what is the problem? My capacity to walk brings with it distinct subclasses of what is possible with respect to it. I know simply in possessing it that taking steps while remaining upright marks a perfection while stumbling and tumbling mark imperfections in it. Yet imagine a sceptic who wonders whether actually I have a capacity to stumble and tumble, with every exercise in which I remain upright is in fact to that extent imperfect. I cannot take such a sceptic seriously. I likewise cannot, though, respond to the sceptic in a way that embarrasses him. Part of knowing something substantive is knowing it despite the logical or even conceptual possibility of its alternatives, at least on views that distinguish conceptual and

metaphysical possibility. Part of knowing something substantive as substantive is knowing that such alternatives are logically or even conceptually possible.

As with my capacity to walk, so with my capacity to act from categorical thought. I cannot take seriously a sceptic who wonders whether I actually have a capacity to act against your good, with every exercise in which I act for your good being to that extent imperfect. Just as my understanding of my capacity to walk includes understanding walking as perfection and tripping as imperfection, so my understanding of my capacity to act from categorical thought includes understanding acting for your good as perfection and acting against it as imperfection. The temptations to deny this knowledge in the latter case might be much greater than the temptations to deny in the former, at least given the structure of our social, political, and economic institutions. Unlike Oscar Wilde, though, I can resist temptation, itself an exercise of my capacity to act that reveals my ability to act from categorical thought and my moral character.

One way to put this point is to say that my reflective knowledge of the principle of my will is in a crucial sense ethical, not epistemological. Such knowledge is not practical knowledge. Still, this reflective knowledge reveals and expresses the ethical character of the agent in question. After all, a being with an awful upbringing in a competitive environment where everyone valorizes traits like ruthlessness, getting ahead of others, and securing yourself at the expense of others might well understand things very differently. He might run for president on a platform of demagoguery and debasement. This person is disturbing, as are the social conditions that help to produce and profit such a person. The disturbance is ethical, not epistemological. They are someone with strained at best moral relationships, and helping them develop their will through training, instruction, and what not will be difficult, especially if they are

set in their ways. Like all such development, crucially parts of this conversion will not be arguments starting from premises that he already accepts and moving towards conclusions with moral content. That is not an upbringing. It is not how development works. He might well fight us along the way, and we might well be in no position to explain ourselves to him without invoking the very understanding of our will towards which we are guiding him and against which he is rebelling.

But so what? The principle of our will establishes normative standards for our development and exercise of it. It is not an externally or internally observable pattern that any agent might simply come across and recognize. It is something you recognize fully only once you develop it. It then characterizes us as the source of our actions and thoughts if we have developed it sufficiently. Yet such development is not an analytical epistemological task. It is the task of a human life, with all of the struggles and dangers, obstacles and illusions, that such a life involves. The principle of my will is not the object of my choice since my choices issue from my will. It also is not such that I can appreciate it no matter my upbringing or course of life simply by examining the concept of a self-conscious action. Reflective knowledge is not so easy. Yet the principle of my will is mine and it establishes the basic standards for my development and exercise all the same, regardless whether I come to understand it and regardless of whether I come to endorse it. The metaphysics establishes that the task is to come to know the principle through reflection after developing and exercising my will as part of developing and exercising it. The task is one of discovery, wherein I find something, learn something, substantive. Yet, in the end, there is nothing regrettable about finding myself left with something I cannot choose to accept or reject, whose authority does not depend on my endorsement or understanding of it, which I well might not have come to know, and of which I might not be able to convince everyone, at least through analytic argumentation. I here

find myself left with myself, with the principle that characterizes the very agency that distinguishes me as a human being. I would not be me were it otherwise even though there is no argument that shows that it could not be otherwise.

The struggle to understand ourselves and the struggle to live an ethical life are part of living a human life. We must be decent to each other because of our nature. We must crusade against solipsism because of what we are. Our confidence in this claim is an ethical confidence born and bred through a human life with others over time, not a philosophical confidence about a conceptual truth. Ethics is a practical subject, in philosophy as it is in life.

BIBLIOGRAPHY

Elizabeth Anscombe [1957], *Intention*, Harvard University Press.

J.L. Austin [1956_a], "A Plea for Excuses", reprinted in J.L. Austin [1961], *Philosophical Papers*, Oxford University Press, 123-52.

J.L. Austin [1956_b], "Ifs and Cans", reprinted in J.L. Austin [1961], *Philosophical Papers*, Oxford University Press, 153-80.

Simon Blackburn [1998], *Ruling Passions: A Theory of Practical Reasoning*, Oxford University Press.

Simon Blackburn [2010], "Truth, Beauty and Goodness", in Russ Shafer-Landau (ed.) [2010], *Oxford Studies in Metaethics* Volume 5, 295-314.

Matthew Boyle [2012], "Essentially Rational Animals", in Günter Abel and James Conant (eds.) [2012], *Rethinking Epistemology* Volume 2, De Gruyter, 395-427.

Matthew Boyle and Douglas Lavin [2010], "Goodness and Desire", in Sergio Tenenbaum (ed.) [2010], *Desire, Practical Reason, and the Good*, Oxford University Press, 161-201.

Michael Bratman [2001], "Two Problems about Human Agency", reprinted in Michael Bratman [2007], *Structures of Agency*, Oxford University Press, 89-105.

Donald Davidson [1963], "Actions, Reasons, and Causes", reprinted in Donald Davidson [1980], *Essays on Actions and Events*, Oxford University Press, 3-20.

Donald Davidson [1973], "Freedom to Act", reprinted in Donald Davidson [1980], *Essays on Actions and Events*, Oxford University Press, 83-102.

Donald Davidson [1974], "Psychology as Philosophy", reprinted in Donald Davidson [1980], *Essays on Actions and Events*, Oxford University Press, 229-238.

Donald Davidson [1982], "Paradoxes of Irrationality", reprinted in Donald Davidson [2004], *Problems of Rationality*, Oxford University Press, 169-88.

Roy Edgley [1969], *Reason in Theory and Practice*, Gregg Revivals.

Stephen Engstrom [2009], *The Form of Practical Knowledge*, Harvard University Press.

David Enoch [2006], "Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Agency", *Philosophical Review* 115, 169-198.

David Enoch [2010], “Shmagency Revisited”, in Michael Brady (ed.), *New Waves in Metaethics*, Palgrave Macmillan, 208-33.

Anton Ford [2011], “Action and Generality”, in Anton Ford, Jennifer Hornsby, and Frederick Stoutland (eds.) [2011], *Essays on Anscombe’s Intention*, Harvard University Press, 76-104.

Anton Ford [2014], “Action and Passion”, *Philosophical Topics* 42 (1), 13-42.

Harry G. Frankfurt [1971], “Freedom of the Will and the Concept of a Person”, reprinted in Harry G. Frankfurt [1988], *The Importance of What We Care About*, Cambridge University Press, 11-25.

Harry G. Frankfurt [1978], “The Problem of Action”, reprinted in Harry G. Frankfurt [1988], *The Importance of What We Care About*, Cambridge University Press, 69-79.

P.T. Geach [1956], “Good and Evil”, *Analysis*, Vol. 17, No. 2, 33-42

Allan Gibbard [2003], *Thinking How To Live*, Harvard University Press.

Allan Gibbard [2008], *Reconciling Our Aims*, Oxford University Press.

Immanuel Kant [1781/7], *Critique of Pure Reason*, trans. and ed. by Paul Guyer and Allen W. Wood [1998], Cambridge University Press.

Immanuel Kant [1785], *Groundwork for the Metaphysics of Morals*, trans. by Mary Gregor, Cambridge University Press. Citations to this work refer to the marginal numbers that mark the academy pagination.

Paul Katsafanas [2013], *Agency and the Foundations of Ethics: Nietzschean Constitutivism*, Oxford University Press.

Christine M. Korsgaard [1983], “Two Distinctions in Goodness”, in Christine M. Korsgaard [1996], *Creating the Kingdom of Ends*, Cambridge University Press, 249-74.

Christine M. Korsgaard [1993], “The Reasons We Can Share: An Attack on the Distinction between Agent-Relative and Agent-Neutral Values”, reprinted in Christine M. Korsgaard [1996_b], *Creating the Kingdom of Ends*, Cambridge University Press, 275-310.

Christine M. Korsgaard [1996_a], *Sources of Normativity*, Cambridge University Press.

Christine M. Korsgaard [1999], “The Myth of Egoism”, reprinted in Christine M. Korsgaard [2008],

The Constitution of Agency, Cambridge University Press, 69-99.

Christine M. Korsgaard [1997], “The Normativity of Instrumental Reason”, reprinted in Korsgaard [2008_b], *The Constitution of Agency*, Oxford University Press, 27-68.

Christine M. Korsgaard [2003] “Realism and Constructivism in Twentieth-Century Moral Philosophy”, reprinted in Korsgaard [2008_b], *The Constitution of Agency*, Oxford University Press, 302-26.

Christine M. Korsgaard [2008_a], “The Activity of Reason”, *The Proceedings of the American Philosophical Association*, Vol. 83, No. 2, 23-43.

Christine M. Korsgaard [2009], *Self-Constitution*, Oxford University Press.

Christine M. Korsgaard [2013], “The Relational Nature of The Good”, in Russ Shafer-Landau (ed.) [2013], *Oxford Studies in Metaethics Volume 8*, Oxford University Press, 1-26.

Christine M. Korsgaard [2014], “On Having a Good”, *Philosophy* 89, 405-29.

Douglas Lavin [2004], “Practical Reason and The Possibility of Error”, *Ethics* 114, 424-57.

Douglas Lavin [2012], “Problems of Intellectualism: Raz on Reason and Its Objects”, *Jurisprudence* 2.2, 367-78.

Douglas Lavin [2013], “Must There Be Basic Action?”, *Nous* 47:2, 273-301.

Douglas Lavin [*ms.*], “Forms of Agency”.

John McDowell [1995], “Might There be External Reasons?”, reprinted in John McDowell [1998], *Mind, Value, and Reality*, Oxford University Press, 95-111.

G.E. Moore [1903], *Principia Ethica revised edition*, edited by Thomas Baldwin, Cambridge University Press.

Thomas Nagel [1970], *The Possibility of Altruism*, Canto.

Brian O’Shaughnessy [1980/2008] *The Will Vol. 1&2*, Cambridge University Press.

Derek Parfit [2011], *On What Matters*, Oxford University Press.

Plato, *Phaedo*, trans. by G.M.A Grube, in John M. Cooper and D.S. Hutchinson (eds.) [1997], *Plato Complete Works*, Hackett.

Peter Railton [1999], “Normative Force and Normative Freedom: Kant and Hume, but not Kant *versus* Hume”, *Ratio* 12(4), 320-53.

Joseph Raz [1998], “Incommensurability and Agency”, in Raz [1999b], *Engaging Reason*, Oxford University Press, 46-66.

Joseph Raz [2011], *From Normativity to Responsibility*, Oxford University Press.

Andrews Reath [1994], “Legislating the Moral Law”, reprinted in Andrews Reath [2006_a], *Agency and Autonomy in Kant’s Moral Theory*, Oxford University Press, 92-120.

Andrews Reath [2006_b], “Autonomy of the Will as the Foundation of Morality” in Andrews Reath [2006_a], *Agency and Autonomy in Kant’s Moral Theory*, Oxford University Press, 121-72.

T.M. Scanlon [2004], “Reasons: A Puzzling Duality?”, in R. Jay Wallace, Phillip Petit, Samuel Scheffler, and Michael Smith (eds.) [2004], *Reason and Value: Themes From the Philosophy of Joseph Raz*, Oxford University Press, 231-46.

T.M. Scanlon [2014], *Being Realistic about Reasons*, Oxford University Press.

Arthur Schopenhauer [1839], *The Two Fundamental Problems of Ethics*, trans. and ed. by Christopher Janaway [2009], Cambridge University Press. Citations to this work refer to the marginal numbers that mark the original pagination.

Kieran Setiya [2005], “Review of *Agency and Answerability: Selected Essays*”, *Mind*, Vol. 114, 786-91.

Kieran Setiya [2007], *Reasons without Rationalism*, Princeton University Press.

Michael Smith [2013], “A Constitutivist Theory of Reasons: Its Promise and Parts”, in *LEAP: Law, Ethics, and Philosophy* (1), 9-30.

Sharon Street [2008], “Constructivism about Reason”, in Russ Shafer-Landau (ed.) [2008], *Oxford Studies in Metaethics Vol. 3*, Oxford University Press, 207-46.

Sharon Street [2010], “What is Constructivism in Ethics and Metaethics?”, *Philosophy Compass* 5, 363-84.

Sharon Street [2012], “Coming to Terms with Contingency: Humean Constructivism about Practical Reason”, in James Lenman and Yonatan Shemmer (eds.) [2012], *Constructivism in Practical Philosophy*,

Oxford University Press, 40-59.

Michael Thompson [2004_a], “Apprehending Human Form”, in Anthony O’Hear (ed.) [2004], *Modern Moral Philosophy*, Cambridge University Press, 47-74.

Michael Thompson [2004_b], “What is it to Wrong Someone? A Puzzle about Justice”, in R. Jay Wallace, Phillip Petit, Samuel Scheffler, and Michael Smith (eds.) [2004], *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, Oxford University Press, 333-84.

Michael Thompson [2008], *Life and Action*, Harvard University Press.

Judith Jarvis Thomson [2008], *Normativity*, Open Court.

J. David Velleman [1992], “What Happens When Someone Acts?”, reprinted in J. David Velleman [2000_a], *The Possibility of Practical Reason*, Oxford University Press, 123-43.

J. David Velleman [2000_b], “Introduction”, in J. David Velleman [2000_a], *The Possibility of Practical Reason*, Oxford University Press, 1-31.

J. David Velleman [2009], *How We Get Along*, Cambridge University Press.

Barbara Vetter [2015], *Potentiality*, Oxford University Press.

R. Jay Wallace [1999], “Three Conceptions of Rational Agency”, in R. Jay Wallace [2006], *Normativity and The Will*, Oxford University Press, 43-62.

R. Jay Wallace [2001], “Normativity, Commitment, and Instrumental Reason”, in R. Jay Wallace [2006], *Normativity and The Will*, Oxford University Press, 82-120.

Gary Watson [1975], “Free Agency”, reprinted in Gary Watson [2004], *Agency and Answerability*, Oxford University Press, 13-32.

Gary Watson [2003], “The Work of the Will”, reprinted in Gary Watson [2004], *Agency and Answerability*, Oxford University Press, 123-57.

Bernard Williams [1980], “Internal and External Reasons”, reprinted in Bernard Williams [1981], *Moral Luck*, Cambridge University Press, 101-13.

Ludwig Wittgenstein [1939], “Lectures on Freedom of The Will, notes by Yorick Smythies”, published in Ludwig Wittgenstein [1993] *Philosophical Occasion: 1912-1951*, James Klagge and Alfred Nordmann (eds.), Hackett Publishing Company, inc., 427-44.