PERSONAL IDEALS AND RATIONALLY IMPOTENT DESIRES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
The Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Regan Lance Reitsma, B.A.

*****

Dissertation Committee:
Professor Donald Hubin, Advisor
Professor Justin D’Arms
Professor Daniel Farrell

Approved by

Advisor
Philosophy Graduate Program
ABSTRACT

A practical reason, in my terminology, is a consideration in favor of doing something, a normative entity that ought to be weighed in when deciding what to do. A neo-Humean subjectivist about practical reasons thinks that there is, in the very least, a standard connection between having a basic, unmotivated desire and having a practical reason to take the means to its satisfaction. Harry Frankfurt, a neo-Humean, believes that this standard connection is sometimes severed. In some instances, an agent has a basic desire, recognizes one or more adequate means, and yet has no practical reason whatsoever to take any of these means because the desire itself “does not deserve a voice” in practical deliberation. Call a desire that is not able to generate practical reasons a “rationally impotent desire.” Is Frankfurt correct? Are there rationally impotent basic desires? And is the neo-Humean able to explain how a basic desire is able to be rendered rationally impotent? I argue that there are rationally impotent basic desires; that the neo-Humean is able to account for them, by appealing to volitional norms contained in personal ideals; and that my subjectivist account of rationally impotent basic desires is stronger and more plausible than the proposals made by Kant and the neo-Kantian Christine Korsgaard.
Dedicated to *myn geweldich heit* and *mem, sportyf Duko, witkop* Annemaïke, *krollekop* Amarinske, and *myn geduldich wyfke*. 
ACKNOWLEDGMENTS

I wish to thank my advisor, Don Hubin, for his personal and intellectual support, his close reading of drafts, his wonderfully clever counterexamples and thought experiments, and for the many large and small ways that he made me a better thinker.

I also thank Justin D’Arms and Dan Farrell for stimulating questions and intellectual challenges. I am grateful to Edwin England for his friendship and the many, many late-night philosophizings. My greatest debt is to my wife; without her patience, this project wouldn’t have been completed.
VITA

April 14, 1972 . . . . . . . . . . . . . . . . . . . . . . . . . Born – Sunnyside, Washington

1994 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . B.A. Philosophy/Classics, Calvin College, Honors


2003 – 2004 . . . . . . . . . . . . . . . . . . . . . . . . . . .  Assistant Professor, Lynchburg College

2004 – present . . . . . . . . . . . . . . . . . . . . . . . . . . Assistant Professor, King’s College

PUBLICATIONS

Research Publication


FIELDS OF STUDY

Major Field: Philosophy
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>v</td>
</tr>
<tr>
<td>1. Introduction: A Perplexity and a Modest Proposal</td>
<td>1</td>
</tr>
<tr>
<td>2. The Anti-Subjectivist Challenge</td>
<td>31</td>
</tr>
<tr>
<td>3. Why It’s Better to Be a Discriminating Subjectivist</td>
<td>77</td>
</tr>
<tr>
<td>5. Kant and the Categorical Rejection of Desire</td>
<td>185</td>
</tr>
<tr>
<td>6. Korsgaard and Practical Identity</td>
<td>264</td>
</tr>
<tr>
<td>7. Conclusion</td>
<td>314</td>
</tr>
<tr>
<td>Bibliography</td>
<td>316</td>
</tr>
</tbody>
</table>
I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn’t dominate me and now I have a problem. Shall I act? Is this desire really a reason to act? (Korsgaard, Sources, 93)

I don’t think odd whims—the passing urge to stick my finger into a gooey substance, or the vertiginous urge to fling myself from a balcony—are considerations in favor of doing those things at all. (Cohon, “Roots,” 63)

A Perplexity

There is a puzzle, as yet unresolved, within Harry Frankfurt’s subtle and otherwise significantly-developed theory of practical reasons.

This puzzle has to do with long-standing disputes between the Humean subjectivist, Kantian constructivist, and (Aristotelian) realist traditions about the ultimate source of practical reasons and the rational criticism of desire.¹ A reader who first confronts the puzzle in Frankfurt’s writing might express it in the following (accurate but sketchy) way: Frankfurt is a subjectivist about practical reasons, in the Humean tradition, who says some very un-Humean-sounding things, a boisterously polemical anti-Kantian and anti-realist who says some very Kantian- and realist-sounding things.

¹ Garrett Cullity and Beris Gaut provide a helpful discussion of these “three poles” in contemporary theorizing about practical reasons (Ethics, 1-27).
Let’s understand a practical reason to be a consideration in favor of doing something. It is a proper input into practical deliberation, a normative entity that ought to be counted in the process of deciding what to do. I will regard a practical reason as a “pro tanto” consideration: it does supply a positive consideration in favor of doing something, one that ought to be weighed against whatever other practical reasons an agent happens to have, though it could conceivably be outweighed by one or more of them. For example, you might have a practical reason to take the afternoon off, but a weightier practical reason to put your nose to the grindstone.

Subjectivism is, among other things, a theory about the ultimate source of practical reasons. It claims that an agent’s practical reasons—all of them—ultimately derive from among the “elements” of “his subjective motivational set, S”—or, put more commonsensically, from among his own desires. To give a stock subjectivist example, your thirst—your basic, unmotivated desire for something to drink—gives you a practical reason to take the means to

---

2 Cullity and Gaut use the term ‘normative reason’ instead of ‘practical reason’ (Ethics). On another note, I agree with Cullity and Gaut that the concept of a practical reason is a fundamental and indispensable aspect of a normative theory of practical reasoning. Any plausible theory will include the concept of a normative consideration in favor of doing something.

3 I don’t mean to deny that there are also considerations in favor of desiring, feeling, or being a certain way. I think there are. I will regard practical reasons, though, as considerations for acting.

4 In my terminology, a practical reason that ought to be acted upon, all relevant considerations weighed in, is a ‘final’ practical reason. One way of making sense of the claim that a pro tanto consideration does have normative authority even though it might be outweighed is to recognize that if it were the only consideration relevant to a particular instance of practical deliberation, it would be a final practical reason.

5 The phrasing is Bernard Williams’ (Moral Luck).

6 A motivated desire, following Donald Hubin’s terminology, is a desire that is generated by another desire, along with beliefs. For instance, if you desire health, come to believe that drinking water is healthy, and subsequently come to desire to drink water as a means, your desire to drink water is a motivated desire. A basic, unmotivated desire is a desire that is not generated from another desire in this way. It is, in this sense, an “original existence” (“Irrational Desires”).


satisfy it, walking to the drinking fountain or to the soda machine, say.\textsuperscript{7} We might put the subjectivist’s defining idea in this way. (From this point, I often will.) In the stock example, your thirst is “rationally potent”: it generates practical reasons. The desire, along with facts relevant to its satisfaction, is the source of practical reasons for you.\textsuperscript{8}

Frankfurt is a subjectivist. He thinks there are rationally potent basic desires, such as your thirst; he never deviates from the theory-defining thesis about the ultimate source of practical reasons: all of your practical reasons derive, ultimately, from among your own subjective, contingent desires; and he thinks that there is, in the very least, a standard connection between having a basic desire and having a practical reason to take the means to satisfy it.\textsuperscript{9} He also explicitly rejects the central tenets of both Kantianism and realism. (Evidence: Frankfurt calls the realist and Kantian programs, each of them, a “will o’ the wisp,” motivated by a “pan-rationalist fantasy” \textit{[Reasons of Love}, 28].\textsuperscript{10} Them’s fightin’ words, meant to accuse these giants in the history of philosophical reflection of having feet made of clay.) Even more to the heart of

\textsuperscript{7} To clarify, I am thinking of thirst, in this example, as an appetite, a species of desire, not as a phenomenological state—a feeling of dryness in the throat and on the tongue—or a physiological state—the body’s biological need for replenishing liquid.

\textsuperscript{8} To be clear, the term ‘satisfaction’ is not a psychological notion. A desire’s satisfaction does not entail that the agent’s consciousness has been affected thereby. If a person wants his family to visit his grave once a year after he is dead, and his family does, his desire has been satisfied in the relevant sense. A desire’s satisfaction entails only that what the agent wished for has come to be the case.

\textsuperscript{9} Frankfurt describes this connection as the “default” condition (“Taking Ourselves Seriously,” 8).

\textsuperscript{10} Here’s the quotation: “We need to understand . . . that the ambition to provide an exhaustively rational warrant for the way in which we are to conduct our lives is misconceived. The pan-rationalist fantasy of demonstrating—from the ground up—how we have most reason to live is incoherent and must be abandoned . . . If we are to resolve our difficulties and hesitations in settling upon a way to live, what we need most fundamentally is not reasons or proofs. It is clarity and confidence. Coping with our troubled and restless uncertainty about how to live does not require us to discover what way of living can be justified by definitive argument. Rather, it requires us simply to understand what it is that we ourselves really care about, and to be decisively and robustly confident in caring about it.” The influence of Descartes’ \textit{Meditations} is evident.
things, to understand Frankfurt’s larger project, I submit, is to understand that Frankfurt is a
subjectivist with an agenda to refine, and so to strengthen, the theory.

What’s striking is that, for all the depth and width of his subjectivism, Frankfurt commits
himself—repeatedly and emphatically—to several ideas uncharacteristic of Humean subjectivists.

To begin with, Frankfurt believes in the existence of basic desires—motivational states,
“elements of S”—that do not generate any practical reasons whatsoever. Though Frankfurt
accepts, in accordance with subjectivism, that there is a standard connection between having a
basic desire and having a practical reason to satisfy it, he thinks that this regular link is, in some
instances, severed. In such cases, there is a basic desire, an adequate means to its satisfaction,
and yet no practical reason to take this or any means because the desire, in Frankfurt’s view, does
not “deserve a voice.” The basic desire is satisfiable but fails, even so, to create any pro tanto
consideration that ought to be weighed into the agent’s decision about what to do. Frankfurt
accepts, we might say, “the rational impotence thesis,” the claim that there are “rationally
impotent” basic desires.

---

11 Frankfurt doesn’t explicitly specify that the motivational states that are rationally impotent are, in my
sense, basic; but as I interpret Frankfurt, this is what he has in mind.

12 The standard connection is facilitated by an instrumental principle that transmits normativity from a
rationally potent basic desire to the means to its satisfaction. A commitment to such a principle is also a
fundamental aspect of any subjectivist theory, as I’ll discuss at greater length in chapter 2.

13 Frankfurt doesn’t put the point in this way. His tendency has been to make a distinction between two
types of conflict between desires. Conflict of the first kind occurs when two desires struggle to determine
the agent’s course of action, competing with each other for priority. The resolution of this type of conflict
requires nothing more than that the agent establish an order between them. Conflict of the second kind is
“deeper”; its resolution “involves a radical separation of the competing desires, one of which is not merely
assigned a relatively less favored position but extruded entirely as an outlaw” (“Identification and
Wholeheartedness,” 170).

14 Frankfurt makes a generalization: rationally impotent basic desires are, statistically speaking, relatively
uncommon; most of the vast number of desires that well up in the common run of life do not call for
Frankfurt’s canonical example of a rationally impotent basic desire is the *unwilling* addict’s present (and very intense) desire to take heroin. More recently, he has proposed another: Frankfurt imagines finding in himself a murderous impulse against his beloved son. Frankfurt stipulates that this distressing mental state is not merely desire-like, sharing some but not the defining features of a desire; the lethal impulse is, as he puts it, a “real” desire. And he is adamant that it is rationally impotent; he calls it “preposterous” to think that the desire gives him a positive reason to take the adequate means to satisfying it: pulling the trigger of the gun that happens to be in his hand (“Taking Ourselves Seriously,” 10-13).

This raises the first in a list of puzzling questions about Frankfurt’s subjectivist program. Don’t subjectivists, following David Hume, commonly describe basic, unmotivated desires as “original existences” (*Enquiry*, 293) and claim that, as such, they do not admit of non-instrumental, rational criticism? And yet Frankfurt embraces the rational impotence thesis. Can a thorough-going Humean subjectivist do that?

There’s more. Frankfurt speaks—and has spoken for thirty years—of the will’s ability to “disenfranchise”: to engage in non-instrumental, “categorical” rejection of basic, unmotivated desires. For example, in constructing the example of the unwilling addict, Frankfurt indicates that the addict *determines* that his intense urge to take heroin is unworthy of a voice in practical deliberation, and the addict subsequently “extrudes” the impulse from the context of decision-disenfranchisement. As he nicely puts it, most “attach to a moving principle within” the agent. That said, he thinks rationally impotent basic desires exist (“Taking Ourselves Seriously,” 8).

---

15 See “Freedom of the Will and the Concept of a Person” and “Identification and Externality.”

16 This example is a remake of one of Gary Watson’s well-known examples: a loving mother who, while bathing her much beloved, bawling infant, feels an urge to drown him (“Free Agency”).
making by an act of will.\textsuperscript{17} Apparently, in Frankfurt’s view this act of will plays a role, some role or other, in rendering the impulse undeserving of a role, any role whatsoever, in decision-making. Frankfurt goes on to remark that if the extruded desire is irresistible and—despite the will’s edict—brings the agent all the way to action, it is a “tyrant” that has usurped the will’s rightful authority to guide action (“Taking Ourselves Seriously,” 10-14). The desire has significant causal power but no right to rule.

In saying this, Frankfurt shares both a political metaphor and its corresponding language with Kant. It is Kant who is best known for making a distinction between the will in its judgment seat, on the one hand, and clamorous desires attempting to gain a hearing before it, on the other. It is Kant who goes to great pains to attempt to establish both the will’s capacity and its authority to govern: to legislate against the claims made by unruly desires and to issue “inexorable commands without promising anything to the inclinations” (Foundations, 405).\textsuperscript{18} And when Frankfurt describes the addict’s impulse to take heroin as a “tyrant”—and as an “alien” and

\textsuperscript{17} David Velleman objects to Frankfurt’s theory that there are “external” desires, desires that are properly extruded from the agent’s identity, on the grounds that it endorses a “common defensive strategy” that is mentally unhealthy (“Identification and Identity”). To extrude an objectionable desire, Velleman thinks, is an act of wishful thinking, aimed at ignoring parts of ourselves we don’t happen to like. As I construe things, to extrude a desire is to treat it as rationally impotent, to deny it reason-giving status. The extrusion of a desire does not imply that the agent is ignoring the desire or that he has deluded himself into thinking the desire does not really exist. On the contrary, an unwilling addict might be all too familiar with his addictive impulses, even weary of them; and far from ignoring his addiction, he could adopt a personal policy to keep a wary eye out for any manifestations of it. There is nothing self-delusional or mentally unhealthy about this. For a similar set of objections to Frankfurt’s remarks about “the self,” see Irving Thalberg (“Hierarchical Analyses of Unfree Action”). Frankfurt is aware of the possibility that agents might engage in the type of “defensive strategy” Velleman is pointing to; he says, “facts about ourselves are often hard to take . . . we find them irreconcilable with what we want to believe. We hide from the truth, it seems clear, because it conflicts with our self-love” (“Faintest Passion,” 95). These “hidings” simply aren’t what Frankfurt means to be talking about when he discusses the “extrusion” of a desire from the context of decision-making.

\textsuperscript{18} The quotation is from standard Academy pagination 405, not Beck’s text.
“external” influence—that has usurped the will’s due authority, he is on the brink of calling it a “heteronomous master,” a phrase made famous by Kant.

Can Frankfurt have these claims, so reminiscent of Kant, without having to take on the central tenets of Kantianism—significant elements of the “pan-rationalist fantasy”—that he explicitly rejects? Or to ask the same question from the opposite direction: isn’t the castigation of categorical norms of practical reasoning a common pastime among theorists in the Humean tradition? And yet Frankfurt eschews it. Frankfurt’s reader might wonder: can a consistent, clear-headed intellectual heir of David Hume believe in such seemingly “deep” rational criticism of desire and a will that has the right to disenfranchise?

A Modest Proposal

I think so, and I intend to explain how—at least one way how. I will argue for the existence of “norms of rational impotence” with subjectivist groundings. People care about some very complex things. My proposal, stated in a few words, is that there are cases in which an agent ought, as a matter of practical rationality and personal autonomy, to disenfranchise a basic desire in virtue of a personal ideal he cares about.

The easy part of my proposal is to describe what a norm of rational impotence is, at least functionally. Divide the process of practical deliberation into two stages: practical reasons are

---

19 Barbara Herman thinks not. She argues that anyone who wants to account for what Frankfurt does—a theory of active, rational agency—should pay attention to “the compelling attraction of following Kant on these matters.” She thinks that to “succeed with an account of the will,” it is necessary to appeal to “a robust idea of practical reasoning” (“Bootstrapping,” 254).

20 I will mention another proposal in chapter 3, though I won’t spend much time making a case for it.

21 There is a distinction between “objective” and “subjective” practical reasons. As proof, consider the following case: a patient receives mistaken advice from a well-reputed doctor about what medicine will
first collected, then subsequently weighed or otherwise compared, in an attempt to determine what ought to be done, all things considered. A norm of rational impotence is relevant to the first stage; it is a rule for the collection of practical reasons, a rule that requires an agent to treat certain basic desires as rationally impotent. In my subjectivist theory, the standard connection between having a basic desire and having a practical reason to take the means to its satisfaction is defeasible, and a norm of rational impotence operates as a defeater. It strips a basic desire of the capacity to generate practical reasons and thereby short circuits the standard process by which practical reasons are created.

The trickier part is to articulate what a personal ideal is and how such a thing derives the practical authority to make demands about how an agent is to collect practical reasons. One remark, to start things off. I regard a personal ideal as a normative standard that generates substantive practical advice, potentially in answer to a wide array of practical questions about how to behave, how to weigh practical reasons, how to feel, what character traits to foster, and—

cure his illness. The patient, in one sense, does not have a reason to take the medicine: it won’t cure his illness. Given the states of affairs he really cares about and the facts relevant to bringing them about, it is not advisable for him to follow the doctor’s advice. This is to say that the patient does not have an object practical reason, what I call ‘a practical reason of rational advice’, to take the medicine. At the same time, given that the patient has a justified belief that the medicine will cure his illness, it seems correct to say that he has a reason to take it. He would be rational to. Given that an agent might have warranted, but false beliefs about what he really cares about, or about the means to his desired ends, he might well perform rationally even if he does not act upon what is, objectively speaking, the final ‘ought’ of rational advice. A consideration that an agent is warranted in treating as a practical reason is, in my terminology, a practical reason of rational virtue. This distinction reflects that a theory of practical rationality concerns both the provision of rational advice and the rational appraisal of actions. Take me to be speaking of objective practical reasons, practical reasons of rational advice, unless I specify that I am not. In chapter 3, I will discuss the relevance of this distinction to my discussion of the rational impotence thesis.

I don’t mean to say that these stages are always separated in the agent’s mind and done, stepwise, in the order I mention. My point is merely that it is natural, when we step back from the process of practical deliberation, to describe it as having two aspects: the collection and the comparison of practical reasons. Collecting is different than weighing.

Michael Bratman discusses treating a desire as reason-giving, and of decisions not to (“Identification, Decision, and Treating as a Reason,” and “Hierarchy, Circularity, and Double Reduction”).
most saliently for my purposes—how to collect practical reasons. Some personal ideals, this is to say, are very complex. This said, the term ‘personal ideal’, I admit, is more suggestive than clear, and part of my task will be to clarify what I do and don’t mean by it, as well as to indicate how a personal ideal is able, consistent with subjectivism, to gain the normative authority to guide an agent’s behavior and deliberation.

For the time being, let’s set aside this significant undertaking, to get at my central idea in a loose, intuitive way.

A Dialectic

Working through an example might be a good place to start. It will make clear what the puzzle—the tension in Frankfurt’s theorizing—is; why it seems difficult to see how a subjectivist could resolve it; and—most importantly—how an appeal to personal ideals and their associated norms of rational impotence helps.

For the sake of clarity, let’s recognize that the intuitions that drive Frankfurt to the rational impotence thesis raise substantive questions about the rational criticism of basic desires. I have remarked that Frankfurt says some very un-Humean sounding things. But I don’t mean to give the impression that the tension in his account is merely terminological. It is sensible, of course, to ask what Frankfurt means when he says that some desires are to be rejected “categorically,” and what subjectivists mean when they express profound doubts about the existence of “categorical imperatives.” I’ll attempt to clarify these remarks, in chapter 2. But talking through the following example should make manifestly clear that the rational impotence thesis leads us into a substantive debate about the relative merits of competing schools of thought.
Think of Frankfurt’s most recent (purported) example of a rationally impotent basic desire: his violent impulse against his beloved son. Frankfurt is adamant that this murderous urge ought to be disenfranchised. Frankfurt doesn’t, though, make clear what prompts him, in his hypothetical example, to reject the impulse categorically, to give it no voice whatsoever. To be charitable, Frankfurt doesn’t claim to, or even intend to. The example is merely supposed to press us to accept the rational impotence thesis, not to provide an account of what makes the impulse rationally impotent or what licenses the will to disenfranchise it. He is directing our attention to a particular desire and making a claim about it, not analyzing it.

To Frankfurt’s credit, his intuition that some basic desires ought to be disenfranchised does seem, in the very least, initially plausible: people find themselves impelled to do some very strange things, and some things far more awful than strange. It is not hard to imagine that many a parent would share Frankfurt’s thought that his murderous impulse, if it is clamoring for attention, should not be merely “muffled” but “silenced”: its demand (“Kill the boy, kill the boy”) should not be ranked very low in a list of priorities, but “extruded” from the context of decision-making. That said, the example naturally raises the question how a basic desire is ruled out, and I want to think about how to answer it. That’s a central goal of my larger project.

First thing, let’s lay out onto the table what is patently clear. Frankfurt does not have a final practical reason to pull the trigger of the gun in his hand. The case against the violent urge, taken as a candidate for satisfaction, is ridiculously over-determined. On very plausible assumptions about what Frankfurt cares about, there are myriad considerations that speak against his killing his son; to name only a few: since he doesn’t want to go jail; since he doesn’t want to feel the shame and the guilt that would overwhelm him if he were to commit murder; since he doesn’t want to meet face-to-face, after the event, with a wife who is enraged, baffled, and
grieving desperately. And from Frankfurt’s own perspective, the most important voice speaking against pulling the trigger is his love for his children. He wants them to flourish; killing his son won’t, to say the least, serve this desired end.

Frankfurt’s imaginary case is a philosopher’s example, but not one of the highly contrived and colorful thought experiments involving a careening trolley that will kill three or four of Frankfurt’s other children if Frankfurt himself doesn’t shoot this one. And Frankfurt’s son isn’t on the verge of being maliciously tortured and brutally killed by a sadistic psychosurgeon, such that he would be better off dead by a quick and relatively painless gun shot. Frankfurt’s pulling the trigger would rob his son, we are to presume, of a valuable future. In summary, Frankfurt means for his example to point to a perfectly ordinary parent, with an (otherwise) very healthy relationship with his son. The final ‘ought’ of rational advice, all considerations in, is ‘Frankfurt, you shouldn’t pull the trigger’. That’s meant to be a no-brainer.

The unusual element in the thought experiment, the thing that is supposed to stick out in our minds, is the distressing urge. The example is supposed to make us consider the urge’s normative status. Frankfurt wants to press his reader to accept the rational impotence thesis on the strength of thinking about it: “The fact that shooting him is likely to kill him gives me no reason at all to shoot him, even though it is true that I have a desire to kill him and shooting him might do the trick” (emphasis his). Frankfurt is convinced, and wants to convince his reader, that there are rationally impotent basic desires; that much—about Frankfurt’s example and the conclusion he wants us to draw—is clear.

The perplexity arises because it’s not at all clear what is supposed to do the normative work of ruling out the impulse categorically. None of the considerations against pulling the trigger mentioned so far, not even the several that I’ve added to the thought experiment, are able
to do that. They operate as instrumental considerations, all of them. It is because Frankfurt wants, in a (more) basic way, freedom from incarceration, from the wrath of his wife, and from the loss of his son that he ought to hold off on the trigger. Holding his finger steady is the best way—the most effective instrument—to save Frankfurt and his family a lot of loss and heartache.

What is needed for the categorical rejection of a basic desire, it seems, is something with a different “shape” than your “average desire”—something such as a norm or a rule or a set of criteria, something with a deontic character, which a basic desire could conceivably violate, or perhaps a procedure or a test that a basic desire could be put through and conceivably fail. The standard subjectivist apparatus—identify basic desires, seek adequate means—doesn’t seem to supply what is needed. A desire simply doesn’t seem to be the kind of thing that is able to “rule out” another desire. Not even a desire that is of central importance to an agent: neither an appeal to Frankfurt’s desire for his son’s welfare nor to his desire to stay out of hot water seem able; they give him instrumental reasons not to act upon the impulse, not non-instrumental reasons to reject it wholesale. What, then, could do this normative work? More to the point, what could a subjectivist appeal to, to do this work?

Anti-subjectivists provide proposals. Realists and Kantians posit extra-subjectivist practical reasons, or extra-subjectivist constraints on practical reasons. They might suggest, respectively, that Frankfurt’s violent impulse is rationally impotent because it fails to track

---

24 This is an old objection to Frankfurt. Watson made a similar objection to Frankfurt’s early theorizing about “internal” and “external” desires, an objection directed at a conception of the will according to which the will consists of an agent’s highest order desires (“Free Agency”). It’s not clear whether Frankfurt ever held this view; but if he did, he has since given it up. As I interpret his current theory, many of the desires that constitute an agent’s will are first-order: Frankfurt’s desire that his children flourish, for instance. That said, Frankfurt hasn’t given up the aspect of this theory that is the purported “trouble-maker.” He did, and continues, to think that the will consists, centrally, of motivational states.
standards of objective worth or because the desire, or its maxim, does not pass a universally valid test of practical reason such as the Categorical Imperative.

Take Kant, for instance. Arguably, Kant develops a non-standard theory of the agency so as to account for the intuition that the will has the authority, as practical reason, to govern, among other things, desires. One feature of governance is getting to decide who gets a say. And so Kant aims to generate an account of the will that captures the idea that it has the capacity to disenfranchise desires. Kant thinks that there is a test of practical reason that an inclination has to pass if it is to play a legitimate role in bringing the agent to action. Kant thinks that an inclination to act is “associated with” a maxim, a principle of choice; and the Categorical Imperative is a test, a rule of rational willing, that takes maxims as proper input. If a desire’s maxim is unable to pass this test, it does not deserve a voice, any voice whatsoever, in practical deliberation. Whether or not a Kant’s theory is ultimately viable, at least it attempts to construe the will in such a way that it has the capacity to “rule out” a desire. The will contains a test of practical reason for the claims of desire.

Frankfurt balks (“Getting Things Right,” 33). He displays the theoretical modesty characteristic of theorists in the subjectivist tradition, and he means to reject these theoretically robust suggestions—these thicker conceptions of “Reason”—when he accuses realism and Kantianism of being motivated by “misconceived ambitions” and a “pan-rationalist fantasy.” (I won’t get into why Frankfurt rejects these ideas just yet; that’s for later. There are, I think,

---

25 See Herman for a defense of this claim (“Bootstrapping”).

26 Several philosophers would propose that Frankfurt’s judgment that he ought to disenfranchise his violent impulse against his son is proper and properly grounded in a judgment of the objective worth of his child. Frankfurt thinks that the conceptual connection between caring about something and judging it important or valuable runs in the opposite direction: “what we love necessarily acquires value for us because we love it. . . my love for [my children] is not at all a response to an evaluation of . . . them” (Reasons of Love, 39).
grounds for skepticism about realism and Kantianism, but absent an account of how a subjectivist theory is able to affirm the rational impotence thesis, someone might ask, skeptical of Frankfurt, “What could rule out basic desires if not something like realist standards of objective worth or Kantian rules of rationally willing?”

A subjectivist, averse to realism and Kantianism and unconvinced by Frankfurt’s intuitions, might try to push Frankfurt in the other direction. She might wonder why subjectivists should bother to appeal to the rational impotence thesis. Even if Frankfurt’s impulse to kill his son were rationally potent, it wouldn’t change what he has most reason to do. The violent impulse, were it weighed in, would be resoundingly defeated. It would be ranked very low in the list of things he has practical reason to do. Not needing the rational impotence thesis to get to the right answer about the final ‘ought’ of rational advice, why argue for it? What’s at stake?

Frankfurt does have options other than realism and Kant. To provide only one, he could entertain the idea that the will has the prerogative—the normative authority—to make whatever rulings it wants: whatever the will chooses is the final word, normatively speaking. If the will chooses to disenfranchise a desire, it thereby becomes rationally impotent for the agent. But that would be a very strange thing for a subjectivist to think. A subjectivist thinks that chains of justification end, when it comes to practical reasoning, in basic motivational states, not in brute choices. (Not to mention, the thesis that the will has absolute practical authority would be a very

Frankfurt consistently resists realist proposals and insists that “Love . . . creates the reasons,” not reasons the love. Frankfurt rejects even Susan Wolf’s (self-described) “modest” realist proposal that what we care about is reason-generating so long as the object of what we care about does not violate “minimal” standards of objective worth (“The True, the Good, and the Lovable”). Frankfurt thinks this theory is guilty of “one thought too many” (“Getting Things Right,” 36-39).

27 There are other conceivable lines of thinking, many of which are not attractive. For instance, Frankfurt could become a skeptic about practical reasons.

28 I will evaluate this type of view in chapter 6.
strange thing for anyone to think: acts of will admit of rational criticism. Say, in a momentary fit of self-loathing, your will legislates that you cut off your right arm. Do you really have to? Or even have reason to?) It would seem that a subjectivist such as Frankfurt is going to have to think that the will’s disenfranchisement of the desire is valid only if it is in service of a subjective, contingent motivational state. But that leads back to the problem raised a moment ago: how do desires rule out desires?

Frankfurt is between a rock and a hard place. He accepts a thesis that he has not yet accounted for, and he wants to account for it by appealing to modest theoretical resources. Critics might remark that either the thesis or the modesty has to go.

Frankfurt, of course, has a fiercely independent streak, for which he is well-regarded. In his philosophical reflections, he tells us, he grapples “with lines of thought” that he confronts “as a human being trying to cope in a modestly systematic manner with the ordinary difficulties of a thoughtful life” (“Preface”). He doesn’t shy away from tension and tends to bring a strikingly independent assessment of the issues marked out by philosophical tradition. This self-reliance has often served Frankfurt very well. New routes of thinking might become available, if you are unwilling to accept the standard terms in which old debates have been conducted. Don’t put new wine in old wineskins. But a strategy that works wonders in one context might fail in the next. When it comes to this new project, fitting the rational impotence thesis together with subjectivism, has one of Frankfurt’s most estimable virtues led him down a fruitless path?

---

29 See Buss and Overton (“Introduction,” xi-xx); also, Stump (“Control,” 33), Bratman (“Hierarchy, Circularity, and Causal Determinism,” 65), and Velleman (“Identification and Identity,” 91).

30 Frankfurt’s independence comes through in his writing. He makes minimal references to the philosophical literature, which makes his writing seem about ideas, not about his place in the philosophical literature. But it also makes the work of a commentator trickier. Where does he fit?
I don’t think so. I think Frankfurt’s ideas are able to fit together. Here’s my proposal, applied to Frankfurt’s example. First, there could well be something at stake, from Frankfurt’s own point of view, other than the life and the welfare of his family. In all likelihood, the most important thing from his perspective is his son’s life and his wife’s mental health. But Frankfurt likely brings an abundance of concerns to the table; among them, he might also care about what kind of person to be.

Frankfurt’s example is thinly described, and I’ve already been thickening it—but in directions that don’t help. Let me build more in, in a direction that will. Let’s say that Frankfurt wants to be a good, or at least a decent, father; he has, it then seems natural to say, a personal ideal of parenting.31 Having confronted his violent impulse, we could easily imagine him saying, “I can’t treat my impulse as reason-giving. Not as a father.” Frankfurt’s judgment that his violent impulse is rationally impotent reflects that he thinks it is inconsistent with his role as a father to treat violent impulses as reason-giving, and he wants to be a good father. In this way, Frankfurt is treating his ideal of fatherhood as normative for him. According to my proposal, norms of rational impotence are elements of some personal ideals. If a person cares about a personal ideal that includes norms of rational impotence, other things being equal, these rules for the collection of practical reasons become valid for his practical deliberation.

Here’s the thing with the right shape. Caring about being a good father is a motivational state with a complex object. To be a good father it is to exemplify certain patterns of behavior and deliberation. In this sense, at least, his desire to be a good father is not an “average” desire.

31 One of my hesitations about the term ‘personal ideal’ is that it has perfectionistic overtones that I don’t intend. I mean for a personal ideal to be a normative standard, and so it makes demands, but it needn’t make impossibly uncompromising demands that are unsympathetic to common human weaknesses. A person is able to treat a personal ideal as normative at the same time that she treats herself with charity in her attempts to exemplify it.
The Groundwork for a (Theoretically Modest) Metaphysics of Personal Ideals

Frankfurt’s general theory of practical rationality and personal autonomy, the product of his three-decade program, serves as the foundation for my proposal, which I construe as a supplement to his subjectivist theory of the will. Frankfurt argues that to make sense of a wide range of intuitions about human agency and personal autonomy, it is necessary to appeal to the concept of an agent’s “practical point of view,” the practical standpoint from which an agent ought to deliberate and make his practical decisions.\(^{32}\) In Frankfurt’s theory, an agent’s practical point of view—his will—centrally consists of various “ruling passions,” and practical rationality and personal autonomy are a matter of serving these passions, with proficiency. Caring about is a complex motivational state, a ruling passion, with the practical authority to guide an agent’s decision-making and behavior.\(^{33}\) (Frankfurt’s subjectivism coming through again.)\(^{34}\)

In Frankfurt’s general theory of practical rationality and personal autonomy, ruling passions have the authority to govern. As Frankfurt puts it, “Love makes demands.”\(^{35}\) At the

---

\(^{32}\) For example, to make sense of the unwilling addict’s sense of personal defeat when his desire for heroin leads him all the way to action, we have to indicate why the addict is to be identified with his desire not to take drugs. In Frankfurt’s most recent work, he appeals to the concept of an agent’s practical point of view to indicate why an agent should identify some courses of action as the things to do, and some desires as worthy of satisfaction.

\(^{33}\) See, also, Fehige’s defense of “the Hearty View” of practical rationality: “Some things are dear to our hearts. To act rationally, I submit, means in essence: to look after these things, as best we can” (“Instrumentalism,” 49).

\(^{34}\) My theory does not give the “principle of self-interest” pride of place. First, an agent might care about things other than, and more than, himself and his own (short- or long-term) self-interest. Second, it is conceivable, though unusual, for an agent not to care about his own welfare at all. For an argument that it is a virtue for a theory of practical rationality not to treat prudence as an intrinsically rational practical standard, see Hubin (“Prudential Reasons”).

\(^{35}\) For Frankfurt, ‘love’ is a technical term. To love someone is to care about her, inescapably. To discover that you cannot help caring about someone (or something) because you are unwilling to oppose the requirements caring about her (it) places on you, when this unwillingness is itself something which you are unwilling to alter is to discover that you love her (it) (“Autonomy, Necessity, and Love”).
heart of a loving parent’s love is her desire for her child to do well, and this ruling passion makes
demands upon her. Behavioral demands: she ought to feed and clothe her child, for instance.
And volitional demands: she ought to treat the fact that a course of action will significantly
benefit her child as, in the very least, a significant reason to do it. A ruling passion might also
call for the adoption of other personal guidelines. For instance, a loving parent who consistently
loses patience with her children, or who finds herself resentful of her children, might decide to
adopt a personal policy “Be more patient” or “Pay more attention to the good things that come
from having children.” These guidelines, as I see it, are properly incorporated into her own
practical point of view. The demands of love are demands upon her. The voice of her love
deserves a privileged seat at the table, with the practical authority—in some cases—to demand
that the agent make significant changes in behavior, volition, and character.

The spirit of my larger argument is that, once we recognize that there are agents who care
about complex personal ideals that include, among other things, norms of rational impotence,
there isn’t any compelling reason to exclude these norms, as a class, from the agent’s practical
point of view. The voice of a valid personal ideal, such as an ideal of fatherhood, speaks, making
demands about how he ought (not) to treat them, about how he ought (not) to feel about his
children, and, I would like to add, what (not) to treat as reason-giving. This voice has the
practical authority to dictate how an agent is to think through his practical decisions in manifold
ways. Why not, then, about how he collects practical reasons?

Practical Reasoning and an Agent’s Normative Self-Conceptions

To locate my proposal within the contemporary literature about practical reasons, various
philosophers have argued that there are practical reasons that are grounded in an agent’s
“practical identity” or his “normative self-conceptions.” For instance, David Copp has argued for “self-grounded” practical reasoning and for “reasons of autonomy” (“Normativity,” 22). Similar to Frankfurt, Copp thinks that some motivational states make up an agent’s “self” in a sense relevant to “self-directed” or autonomous agency. These basic desires have special authority for an agent, and they commonly prompt self-referential remarks. A self-described and personally committed pacifist might say, “I can’t treat war as a practical option,” with a special emphasis upon the “I.” Christine Korsgaard, a contemporary Kantian, thinks that an agent’s normative self-conception, or “practical identity,” is a source of practical advice:

A century ago a European could admonish another to civilized behavior by telling him to act like a Christian. It is still true in many quarters that courage is urged on males by the injunction ‘be a man!’ Duties more obviously connected with social roles are of course enforced in this way. ‘A psychiatrist doesn’t violate the confidence of her patients.’ No ‘ought’ is needed here because the normativity is built right into the role (Sources, 101).

This idea that some practical reasons are connected to normative self-conceptions forms the foundation of my own proposal.

My supplement to this line of thinking is that there are “factors” relevant to practical deliberation that have a different shape than practical reasons. Discussions of practical reasons are replete with (useful) metaphors. A standard way of thinking about a practical reason is a gram weight on a scale: “I have a reason to X, but a weightier reason to Y; so Y is the thing to do.” A norm of rational impotence functions differently. It renders a basic desire incapable of generating practical reasons for an agent. Norms of rational impotence, I submit, are often,

---

36 See, also, Nagel (Possibility), Parfit (Reasons and Persons), and Hubin (“Desires, Whims, and Values”).

37 The relationship between ruling passions and the relative weight of the practical reasons they generate is not simple. A ruling passion generates, in some cases, very weak practical reasons. Even if I care about health deeply, I don’t have a significant practical reason to do each and every action that would benefit it.
though perhaps not invariably, connected to normative self-conceptions. This idea, too, has a precedent in the literature. Korsgaard tends to emphasize this link.

Though my own proposal links practical reasons to an agent’s normative self-conceptions, I don’t want to make the link tighter than it is. Many practical reasons derive from basic desires with no connection to personal ideals. Your thirst generates practical reasons for you even if it is not connected, in any interesting sense, to one of your self-conceptions. And many “innocent whims” are rationally potent. If an agent has any personal ideals that are properly grounded in his ruling passions, they are able to play a role in shaping how he ought to make practical decisions. Frankfurt, the character in my bolstered example, does.

Some Important Questions for Appeals to Personal Ideals

Accounts that link (some) practical reasons to personal ideals raise an array of important questions. These questions call for answers from any theorist—subjectivist, Kantian, realist, or otherwise—who appeals to the concept of a personal ideal. It is in answering these questions that my proposal will begin to be more than a gesture at a theory.

It is often difficult to specify the content of a personal ideal. First, its content might be complex; as mentioned, a personal ideal could include practical advice about behavior, volition, feeling, and character—maybe more. Second, the content of a personal ideal, such as an ideal of

38 For instance, Velleman describes the connection between acting and thoughts about oneself as very tight. He thinks that “the constitutive aim of action” is to “make sense to ourselves” (“What Happens”). I disagree. If I were to think of action as having a single, constitutive aim, I would say that the aim is to bring about states of affairs. Being intelligible to ourselves is only one state of affairs an agent might aim to bring about; we often aim to alter states of affairs outside of our own minds.

39 “I am a person for whom thirst generates practical reasons” is not an interesting sense, I submit.

40 Hubin has the same intuition (“Desires, Whims, and Values”).
parenthood, could be a matter of dispute, intrapersonal or interpersonal: one person thinks that a
good parent spanks, another is repelled by the idea; in some moments, you think a parent ought to
pay for a college education, in others that it is beyond the call. Third, many personal ideals are,
in terms of content, vague or indefinite—to varying degrees. One explanation for the
complexity, disputability, and indeterminacy of personal ideals: they are culturally variable and
change, even within a culture, over the course of time. To what degree is an important question.

Adopting a personal ideal might have practical benefits. A personal ideal that is clear and
determinate serves to orient an agent in normative space, giving him substantive practical advice
about how to govern his own conduct. On the other hand, this practical benefit isn’t guaranteed.
An extraordinarily complex, highly disputed, or unduly vague personal ideal could thrust an agent
into difficult normative questions and disputes. If he can’t seem to settle them, he will not gain
the orientation in normative space a personal ideal promises to provide.

My proposal also raises questions about agents who adopt multiple, conflicting personal
ideals. A person could have competing self-conceptions: in cases in which they conflict, which
should she obey? Also, we could easily imagine an agent with a personal ideal that is harmful, or
otherwise rationally inadvisable. A person without natural athleticism takes on the daily routine
of a jock and suffers under the weight of his (predictable) failures. Wouldn’t it be better, we
might wonder, if he didn’t treat an ideal of athleticism as normative? I agree that there are
instances in which a person is deferential to personal ideals that do not deserve it. I will argue

41 In response to these issues, I subscribe to the view that some practical reasoning consists in specifying
ends, in some cases, in specifying the content of a personal ideal. It is often, as I see it, necessary, if a
person is to live up to a personal ideal, to find ways to make more determinate the very general or vague
practical advice it gives. Conceivably, an agent could, in the process of specifying the content of his
personal ideal, adopt norms of rational impotence, deliberately and self-consciously.

42 I’ll discuss this in chapters 4 and 6.
that an agent has to care “in the right way” about the personal ideal.\textsuperscript{43} That’s weazly, but I’m not giving an analysis yet. I’m simply raising a question that I’ll discuss later.

Another important question is how personal ideals function within practical deliberation. I suspect that the true answer will be several-fold. In some cases, the content of a personal ideal is clear, even codified. A person who wants to be a good psychiatrist will find some of the rules of proper conduct written in the relevant documents published by the American Psychiatric Association, or widely promulgated in discussions with other practitioners: ‘A psychiatrist doesn’t violate the confidence of his patients’. But in some instances, agents operate, I suspect, with a type of mental picture of what an exemplar of the ideal is like and confronts norms of rational impotence in reflection upon, or imagination of, what an exemplar would do. (Think of the bracelets worn by some Christians: “What Would Jesus Do?”) This calls to mind one of Kant’s ideas. Kant thinks that what the fully rational agent would do is what you and I should do. If the fully rational agent would employ a rule of rational willing in his practical deliberation, then you and I should employ it; if the fully rational agent would have a virtue such as self-control, then you and I should have, or at least aspire, to it. The fully rational agent is a hypothetical being, a mental construct, that serves as an example that we are to model our own behavior after. Likewise, a father might have an idea about what a good father is like, and treat it as normative for his behavior, volition, character, and so on.

In such cases, norms of rational impotence are codifications of the activity and traits of exemplars, after whom we might aim to model our own behavior and traits. If the model or

\textsuperscript{43} If, from the perspective of several of an agent’s ruling passions, it is rationally advisable for the agent to treat a personal ideal as normative, and none of his ruling passions speak against it, then it is certainly being cared about “in the right way.”
exemplar would consistently treat a certain basic desire as rationally impotent, we are able to
ascribe a norm of rational impotence to him.

These are good questions, all. As I see it, they are questions that any theory that
appeals to personal ideals has to answer. Given that Korsgaard appeals to the concept of a
personal ideal in her refinement of Kant, and I do in my refinement of Hume, one question is
which theory answers them better. My account, I submit, will handle these cases in a very
plausible and natural way, better than Korsgaard’s Kantian theory.

Agent-Relative Practical Reasons and Agent-Relative Norms

It might be useful for me to mention some of what I don’t intend to defend. I want to
advocate the thesis—which I take to be relatively weak—that there are personal ideals that, at one
and the same time, are sufficiently determinate to guide behavior and practical deliberation,
include norms of rational impotence, and are normative for an agent in virtue of one or more of
his ruling passions.

I accept the rational impotence thesis, and so I agree with the thrust of the remarks,
quoted at the beginning of this chapter, made by Rachel Cohon: there are motivational states—
odd whims, vertiginous urges, and so on—that don’t generate practical reasons. But I want to be
clear that I don’t accept all purported examples of rationally impotent basic desire. I don’t agree,
for instance, with Cohon that the urge to stick a finger in a gooey substance is—at least
generally—rationally impotent. I have a friend, Ryan, with an anti-conventional streak; he
disdares bourgeois standards of cleanliness and openly revels in the disgusting. (To illustrate, he

44 For a thoughtful discussion, see Sen (Reason Before Identity). He also discusses the relevance of
normative self-conceptions to economic theory.
kept a collection of his own nail clippings in a jam jar for many years.) I’m inclined to think Ryan would, and properly should, treat any “gooey urges” he finds in himself as reason-giving. If he were to experience a basic desire to stick his finger in goo, this desire would be rationally potent for Ryan.⁴⁵

According to my subjectivist proposal, the norms that render some basic desires rationally impotent are in a very important sense “agent-relative.” They aren’t realist standards of objective worth or Kantian principles of practical reason, as such. They are norms grounded in the agent’s ruling passions, and they are valid—“normative”—for an agent only if he has the relevant ruling passions.

The Upshot

What does it matter whether I’m right? My proposal has two significant implications, one for subjectivism, another for the long-standing dispute between subjectivism and its opponents.

First, if I am correct, the concept of a norm of rational impotence belongs in our best subjectivist model of agency and practical reasoning, along with the associated concept of a personal ideal; there is useful work for them to do. (Frankfurt himself has gestured at this idea; I want to make good on it.) Embellishing the subjectivist model of agency in these ways, I submit, does not threaten the core of the subjectivist theory. On the contrary, it puts—I will argue—the central insights of the subjectivist tradition into better focus, and it makes its model of normative

---

⁴⁵ The fact that the impulse is fleeting or bizarre is not sufficient to render it rationally impotent. Strange and quirky whims often do generate practical reasons. As I’ll discuss in chapter 2, subjectivists operate, in the very least, with a very inclusive notion of what is able to count as a rationally potent basic desire. If Frankfurt is going to rule out a basic motivational state on the mere grounds that it is fleeting or strange, he will be rejecting a very common and, I think, well-justified aspect of subjectivism.
practical reasoning and autonomous agency more subtle and more attuned to the shape of the
demands made by the ruling passions of both conceivable agents and, I think, real-life, flesh-and-
blood human beings.

Second, the refined subjectivist theory that emerges from my arguments is well-suited—
better-suited than its historical predecessors in the Humean tradition—to stave off a common,
anti-subjectivist line of objection implicit in the literature. Imagine an anti-subjectivist with an
apologetic aim (a realist or a Kantian; at this stage, it doesn’t matter). She wants to convert
subjectivists and fence-sitters to her version of anti-subjectivism, and she needs to find a way to
get this conversion process going, some common ground or some intuitive thoughts to begin to
create suspicions about subjectivism. An anti-subjectivist with this self-appointed task might
begin by sketching an example in which an agent, real or imagined, experiences a bizarre or
unruly or distressing desire, D, and then raise the following argument: “You don’t really think
that D is the source of a practical reason, do you? But subjectivism says that it is. Isn’t this a
serious problem for subjectivism?” (This objection is a subtext in Cohon’s “The Roots of
Reasons.”)

This line of reasoning is a challenge to the Humean subjectivist theory. It accuses the
theory of being insufficiently critical of human desire, of giving a voice to basic desires that don’t
derserve it. If my refined subjectivist theory is correct, this anti-subjectivist argument is unsound;
subjectivism is able to countenance rationally impotent basic desires. What follows is that the
generic intuition that there are rationally impotent basic desires is not, in itself, a reason to suspect
or to reject the Humean subjectivist tradition.
I’m in for a fight, of course. I admit that my theory is unlikely to satisfy some anti-subjectivists; my proposal does not promise to divide up the cases in the way that many realists and Kantians prefer. For instance, I will not claim that basic desires are rendered rationally impotent by realist standards of objective worth. I ground norms for rational impotence in subjective motivational states, ruling passions. Also, I will not argue for the thesis that every immoral basic desire fails, by its very nature, to eventuate in a practical reason, as Kant does. I ground norms of rational impotence in personal ideals, and not every personal ideal is a moral ideal. I ground them in love, and people love some morally abhorrent things. So, if Kant or a latter-day Kantian insists that an adequate theory of practical reason has to capture this thesis, mine won’t satisfy his demand. I’ll leave it to the reader to decide how much stock he wants to give to Kant’s desiderata, or to various other intuitions and aspirations someone might bring to theorizing about practical reasoning.

The implication of my argument for the long-standing dispute is modest, in one sense. It doesn’t prove subjectivism true, or anti-subjectivism false. As mentioned, it implies that the thought that there are satisfiable, basic desires that don’t eventuate in practical reasons is not reason enough to be suspicious of the Humean subjectivist tradition. This is not, though, an insignificant conclusion. The competing traditions are, I think, fighting for contested ground, for various intuitions about practical rationality and autonomous agency, and steady progress in

46 Frankfurt anticipates objections to his theory: “Love is paradigmatically personal. What people love differs, and may conflict. There is often, unfortunately, no way to adjudicate such conflicts. The account of normativity that I have been giving may therefore seem excessively skeptical. Many people are convinced that our final ends and values—most urgently our moral values—must be impregnable authority that is altogether independent of anyone’s personal desires and attitudes. What we should care about, they insist, must be determined by a reality entirely other than ourselves. My account is likely to strike them as radically neglectful of these requirements. They will have the idea that it is unacceptably non-cognitive and relativistic” (“Taking Ourselves Seriously,” 26). Anti-subjectivists are likely to find my general theory as “hopelessly non-cognitive and relativistic.”
accounting for a wider and wider range of the most sensible and compelling intuitions is important progress in this complex dispute.

To resurrect the anti-subjectivist challenge, the proselytizing anti-subjectivist would have to appeal to examples of rational impotence that cannot be accounted for, plausibly, by appealing to norms of rational impotence grounded in personal ideals the agent cares about. It’s beyond the scope of my project to follow out this type of argument.

Why Frankfurt?

A word or two about Frankfurt, whose three decades of theorizing has generated the perplexity I aim to dissolve. I don’t mean to be accusing Frankfurt of a clumsy inconsistency. On the contrary, I think that Frankfurt is pressing in the right direction. And that Frankfurt might seem to be speaking, at times, from competing traditions is altogether unsurprising, given his larger theoretical project. Frankfurt is a restless subjectivist, perturbed by various aspects of the Humean tradition. Subjectivism has been associated, in some minds, with a “minimalist” theory of human psychology and a simple, straightforward belief-desire model of action. Frankfurt is not a minimalist. He makes this vividly clear in the following passage, which also reveals his characteristic, rollicking style:

When philosophers or economists or others attempt to analyze the various structures and strategies of practical reasoning, they generally draw upon a more or less standard but nonetheless rather meager conceptual repertoire. Perhaps the most elementary as well as the most indispensable of these limited resources is the notion of what people want—or, synonymously (at least according to the somewhat procrustean convention that I shall adopt here), what they desire. This notion is rampanty ubiquitous. It is also heavily overburdened, and a bit limp. People routinely deploy it in a number of different roles, to refer to a disparate and unruly assortment of psychic conditions and events (Reasons of Love, 10; emphasis his).
Frankfurt wants to do better. He wants a less meager conceptual repertoire, a more subtle taxonomy of motivation, and he thinks that our best subjectivist model of practical reasoning and agency ought to include, among other things, the concepts of ‘caring about’, ‘loving’, and an agent’s ‘practical point of view’. Much of his work over the past three decades has attempted to clarify these concepts.

Since Frankfurt is a refining subjectivist who wants a theory of personal autonomy, of active rational agency, he is going to speak in ways similar to Kant. They’re attempting to account for similar intuitions, thinking about similar cases.

At the same time, the subjectivist theory is characterized by theoretical modesty. Subjectivists have the firm intention not to appeal to realist standards of objective worth or to a robust Kantian conception of the will. Frankfurt shares this subjectivist characteristic. He has been described—aptly, I think—as one of the “premier modern practitioners” of theoretically abstemious “bootstrapping” (Herman, “Bootstrapping,” 253). Bootstrapping is an argumentative strategy, which Herman describes well:

The attraction of bootstrapping is that you use a little of what you already have to get some place you haven’t been before, but need to go. As a strategy of argument, it is environmentally neutral. No new resources—new entities or capacities—are called for; little of what you start with is wasted. The most compelling bootstrapping arguments occur when the subject is basic—objects; the causally ordered items of experience; the choices of a perduring self—analyzed in the first instance in a theoretically plain way. The resources used in the first explanation are such ordinary items as perceptions and concepts, desires, beliefs, and the like. The bootstrapping move is prompted when further thought about what the basic subject is like reveals complexity, or some vulnerability to skeptical challenge, that outstrips the carrying capacity of the first explanation. The bootstrap is made of the same basic materials, only attached in a different place, so as to provide additional elements of argument. The idea is to provide

47 Frankfurt is not alone. See many of the essays in Millgram (Varieties). Millgram structures his discussion of contemporary theorizing about practical reasoning by beginning with an exceedingly simple, subjectivist theory that accepts an instrumental principle and the claim that practical reasons derive, ultimately and only, from basic desires. He then categorizes various theorists as either complicating or rejecting one or another feature of this simple theory. I think this is a very useful way to go about things.
just enough leverage in the argument to give a better account of the complexity and/or meet the skeptical challenge. Its ontological and theoretical abstemiousness make bootstrapping one of the more elegant modes of philosophical prestidigitation (253).

Frankfurt attempts to bootstrap his way to an account of the human will and to a theory of practical reasoning and personal autonomy.

As mentioned, I myself accept the general contours of Frankfurt’s bootstrapped account of the will. The question for my project is to see whether theoretically modest proposals—more bootstrapping or something similarly humble—can take us one step further. Can modest proposals get us to an adequate account of the rational impotence thesis and the will’s ability and practical authority to categorically reject some basic desires?

According to my proposal, ‘yes’. In one important sense, my proposal does more than appeal to a generic mental state—“the same basic material”—such as a desire—first-, second-, or higher-order—“attached in a different place,” doing an extra bit of work. Broadly speaking, Frankfurt’s theory of the will does this: an agent’s practical point of view consists of complex motivational states with which the agent is “satisfied.” But the norms of rational impotence—the “something with the right type of shape”—that are an essential part of my proposal are not a type of belief or desire. They are rules connected to personal ideals. In my proposal, then, norms of rational impotence enter the agent’s will because the agent desires, in the proper way, an object that is normatively complex. The complexity is a function of the object desired, and it takes a

48 I suppose, conceptually speaking, a desire to follow a norm of rational impotence would also do. But this move smacks of being ad hoc, in part because it is hard to imagine that this is the standard way that norms of rational impotence do enter an agent’s practical point of view. I think, instead, that the norms enter the process of practical reasoning through caring about personal ideals which contain them. Say an agent were to find in himself a desire not to treat certain basic desires as rationally potent. It strikes me as more plausible to interpret this desire as the consequence of something such as a personal ideal he accepts, applied to his own circumstances, then as a basic desire.
sensitive appraisal of the contents of the various personal ideals that people adopt to capture the relevant intuitions.

That’s the basic contours of an answer. It’s time to give a more thorough treatment.
CHAPTER 2
THE ANTI-SUBJECTIVIST CHALLENGE

Each impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action. (Korsgaard, Sources, 91)

When you deliberation, it were as if something were over and above all your desires, something which is you, something which chooses which desire to act on. (100)

Let’s think polemically, in terms of an argument against subjectivism about practical reasons. Doing so will structure my discussion of the rational impotence thesis in a helpful way. In the introductory chapter, I imagined an anti-subjectivist who thinks that subjectivism is insufficiently critical of human desire: more precisely, the subjectivist theory gives, she thinks, a voice in practical deliberation to basic, unmotivated desires that do not deserve it. This anti-subjectivist argument has two premises. The first is the rational impotence thesis, the claim that there are rationally impotent basic desires. The second is the claim that subjectivism is incompatible with the rational impotence thesis. I’ll call this claim “the incompatibility thesis.”

There are two general ways for a subjectivist to respond to the anti-subjectivist challenge. The more common way is to reject the rational impotence thesis; the other way is to reject the incompatibility thesis. Frankfurt, who thinks it is “preposterous” to think that his violent impulse

---

49 An anti-subjectivist could present other challenges to subjectivism. An anti-subjectivist might argue that there are desires that any agent should have—as a matter of practical rationality; subjectivism, problematically, doesn’t entail that every agent should. Or she might argue that there are things an agent should—as a matter of practical rationality—do even if he doesn’t happen to have a desire either to do it or for the behavior’s expected consequences. These challenges simply aren’t the topic of my discussion.
is rationally potent, would take the unconventional strategy, though he has not yet indicated, except in a rudimentary way, how he would go about it. ⁵⁰ I will consider what subjectivists could say against the rational impotence thesis. This tactic isn’t, to my mind, preposterous, but in the end I will take the unconventional strategy, too. I will argue for the existence of norms of rational impotence grounded in personal ideals that an agent cares about.

A Debate

The anti-subjectivist challenge sets up a debate with, at least initially, three participants. Like Frankfurt, I am caught between an anti-subjectivist who thinks I am appealing to intuitions—about the proper disenfranchisement of (some) basic desires—that should lead me to a robust conception of reason and a fellow subjectivist who thinks our shared theory, with its characteristic theoretical modesty, should lead me to give up or explain away the intuitions. Wanting both the rational impotence thesis and the modesty, I will have to battle on two fronts.

As for the first front, I will take up the family dispute against my fellow subjectivist in the third chapter. I will resist the temptation to construe this internecine spat as a debate between “crude” and “sophisticated” versions of subjectivism, mine being the sophisticated version. That would be (not only self-congratulatory but) misguided. Of course, a subjectivism that gives a voice in practical deliberation to every basic desire is, from my perspective, lacking a certain something; and since it treats all basic desires as rationally potent, it is—in this one respect—“indiscriminate.” But such a theory is not, by its very nature, “crude,” and it need not suffer from a “meager conceptual repertoire.” Subjectivism is capable of interesting complexity, as a

⁵⁰It is more accurate to say that Frankfurt has not settled, decisively, on an account. He has gestured at one; see his discussion of “decisive commitment” (“Identification and Wholeheartedness,” 168).
proliferation of work in recent decades has made clear. Accordingly, in chapter 3 I will discuss a number of ways subjectivists could analyze the cases Frankfurt has in mind.

The question relevant to this discussion is whether the subjectivist theory should be refined in a particular way, namely, to include the rational impotence thesis and whatever (novel) conceptual apparatus (if any) is necessary to account for it. I will argue that doing so is a natural extension of several of the best ideas in the subjectivist tradition, an extension which coheres with some of the best recent work done by subjectivists. I think that there is positive evidence for the rational impotence thesis, and that this thesis, properly understood, not only coheres with the subjectivist theory, but strengthens it.

As for the second front, I will also have to argue against anti-subjectivists who either accept the incompatibility thesis or, more ambitiously, make proposals that purport to capture the rational impotence thesis, and the various intuitions that motivate it, on anti-subjectivist grounds. At the end of this chapter, I will argue that the incompatibility thesis is false. For reasons that will emerge, I intend to argue, in the fifth and sixth chapters, against several, positive Kantian proposals: first, Kant’s own argument that the Categorical Imperative is an objective, universally valid rule of rational willing that disenfranchises some basic desires; and second, Christine Korsgaard’s attempt to save Kant’s argument from several standard objections.

Notice that each character in this debate shares at least one claim with the other two characters. I share an underlying theory with my fellow subjectivist and the rational impotence thesis with the anti-subjectivist. My two opponents are also of one mind about at least one claim, the incompatibility thesis, or at least something like it. The thesis is vague; it doesn’t say what type of inconsistency Frankfurt and I allegedly suffer from. Perhaps the claim is that the rational

---

impotence thesis is logically inconsistent with theory-defining elements of subjectivism, or perhaps it is merely that there simply isn’t any plausible way for subjectivism, given the theoretical constraints it self-imposes, to account for it. In any case, in this debate, the anti-subjectivist and my fellow subjectivist both think that, in one way or another, the rational impotence thesis simply doesn’t fit into subjectivism.

Consequently, over the course of my discussion, I will draw three polemical conclusions. The first, directed against both of my opponents, is that it is logically possible for a subjectivist to affirm the rational impotence thesis—it’s not inconsistent with any theory-defining elements; the second, directed against my fellow subjectivist, is that a subjectivism that does affirm this thesis is a stronger theory than one that doesn’t; and the third, directed against Kant and Korsgaard, is that my subjectivist proposal for how to account for the rational impotence thesis is significantly more compelling than their proposals, both of which claim that norms of rational impotence are grounded in something “deeper” than ruling passions.

The Agenda

To this point, I’ve presented a challenge to subjectivism, but it’s schematic and I haven’t done much to motivate it yet. Mostly, I’ve discussed a tension in Frankfurt’s theorizing and set up a three-member debate. In this chapter, I would like to accomplish two goals: to clarify the terms of the debate and to lay out the initial arguments for and against the two premises of the anti-subjectivist challenge. What lines of thought might lead someone, at least initially, to endorse these premises? How might subjectivists respond, at least initially?

It is also necessary to specify, at some point, what conclusion is to be drawn from the premises of the anti-subjectivist challenge if they are true. Is the argument to be taken as a
disproof of subjectivism—a “theory-killer”—or something weaker? Even if the anti-subjectivist challenge constitutes a sound objection to subjectivism, we would have to determine how serious an objection it is, all things considered. In brief, I intend the anti-subjectivist challenge to be a starting point for discussion.

At this stage in the discussion, a cautionary word is in order. We should take care not to read more into the rational impotence thesis than it actually says. The thesis itself leaves a number of important theoretical questions open. For the sake of illustration, imagine for a moment that you do accept the rational impotence thesis, and you want to begin to sketch a theory of normative practical reasoning that accommodates it. The thesis doesn’t indicate what makes a rationally impotent basic desire rationally impotent. There are also questions about the level of importance of the rational impotence thesis. How frequently are basic desires ruled out? Is the rational impotence thesis a “central” thesis, around which a theory of practical reasoning should be built? Or is it a true, but relatively minor or ancillary, thesis? A full account would have to answer these questions.

My account does. I’ve already mentioned what I think the “impotence-making” features (often) are: rationally impotent basic desires are rendered so by norms of rational impotence with subjectivist groundings. As for the other two questions, I don’t think that norms of rational impotence are a necessary element of an agent’s practical point of view; some agents, perhaps, aren’t committed to any. In that sense, norms of rational impotence are not “central” to agency. As for the frequency with which basic desires are rationally impotent, that depends upon how often people care about the personal ideals that contain norms of rational impotence and the frequency with which those people happen to experience desires that violate them.
My discussion will address a wide range of theoretical questions, but it will also be, in one sense, narrow. I will restrict the debate to only a few participants. A general historical study would direct us to various schools of thought from which we could derive proposals for how to account for the rational impotence thesis. For instance, realists of various stripes would make suggestions. A realist with a Platonist bearing might regard a desire as rationally impotent if the object of that desire does not participate (to a sufficient degree) in the Good. If the object of a desire is not (sufficiently) objectively valuable, the desire does not give its bearer a practical reason. Rachel Cohon seems to have sympathy for this type of view (“Roots”), and Susan Wolf has defended a “pluralistic” and “non-elitist” version of this standard (“The True, the Good, and the Loveable”). A realist standard with an Aristotelian flavor would require that a desire be appropriately related to the good where ‘the good’ is understood in terms of human flourishing.\(^52\) Also, Kant’s test of the Categorical Imperative, as I illustrated in the introductory chapter, might be taken to be the relevant standard: a desire is rationally impotent if acting upon the maxim of that desire would implicate its bearer in any of several types of contradiction in the will Kant identifies. There are other conceivable proposals. The divine command theory is generally taken to be a normative ethical theory that states that the substantive content of morality is determined by what God commands and forbids. But we could imagine a theological voluntarist about practical normativity who treats God’s will as the source of the practical advice that is normative for you and me: what God commands you to do is what you should, rationally speaking, do.\(^53\) (Religious believers might well think that sinful desires—“temptations”—are rationally

\(^{52}\) See, for instance, McDowell’s discussion of the “silencing” of desires (“Moral Requirements”).

\(^{53}\) The questions that naturally arise for such a theory mimic the questions that the Euthyphro Dilemma raises: Is God’s will the ultimate arbiter of rational advisability? Does His forbidding a desire make it impotent, or does its being impotent make Him forbid it? If He shapes His willings in accordance with a standard that is independent of His will, what is this standard and how does it have the authority it does?
impotent. In summary, it would not be difficult to conceive of a wide array of anti-subjectivists, allied against subjectivism, who disagree with each other about which desires ought to be disenfranchised and why.

I don’t intend to discuss each conceivable proposal—my discussion won’t be full, in that sense. I’ve chosen my primary dialectical partners. As Cullity and Gaut remark, most contemporary theorizing about practical reasons tends to come from three schools of thought: Humean subjectivism, Kantian constructivism, and (Aristotelian) realism. I will say a few things about realist proposals, but much of my discussion will focus upon subjectivism and Kantianism, which reflects both my commitment to subjectivism and my belief that Kantians are my most significant competition, since they attempt to account for some of the same intuitions—about personal autonomy and about an agent’s construction of a practical point of view—that drive Frankfurt and me to refine the subjectivist theory of normative practical reasoning.

In summary, the anti-subjectivist challenge will structure my discussion of the rational impotence thesis and its theoretical implications, and I aim to generate a stronger, more subtle subjectivist theory in the process of working through this argument. By the end of this chapter, I think it will become clear that the incompatibility thesis is false; a thorough-going subjectivism is able to affirm the rational impotence thesis without giving up anything of central importance to it.

54 For a similar idea, see Terence Penelhum (Christian Ethics and “Human Nature and External Desires”).

55 To my knowledge, there isn’t much precedence for a subjectivism that affirms the rational impotence thesis. Is this evidence for the incompatibility thesis? Very weak evidence, at best. There are other conceivable explanations for this lack of precedence. Hubin, for instance, thinks that Humean subjectivists have often focused their attention on defeating competing theories instead of attempting to complicate their own (“Desires, Whims, and Values”). Also, some subjectivists have been careful to leave open the line of thinking I am taking. Williams’ quotation, which I referenced early in the first chapter, could be interpreted as carefully structured to do this: an agent’s practical reasons derive, he says, from “among the elements” of his “motivational set, S” (Moral Luck). Of course, on any subjectivist view, there are “elements”—derivative desires, based on false means-end beliefs—that do not generate practical reasons. Williams’ gloss, though, leaves open the possibility that there are rationally impotent basic desires.
Common Ground

Before the polemics begin, let’s find some common ground. Participants in this debate ought to agree about the following three ideas.

First, there is a phenomenon, or a set of phenomena, in need of analysis. People do, in some instances, take strongly negative attitudes to their own motivational states. Cohon and Frankfurt provide hypothetical examples in which an agent experiences a peculiar mental state, calls it an “urge” or a “whim” or a “mere desire” or a “temptation,” and subsequently “rejects” or “denounces” it in some way. These examples point to real experiences, common enough in the lives of reflective agents. A complete theory of practical reasoning would indicate how it would analyze—account for, explain away, whatever—these mental states and the corresponding “acts of denunciation.” The first salient point of agreement is merely this: theorists about practical reasoning—whether subjectivist or anti-subjectivist—do share, in the very least, something to think, and so something to disagree, about.⁵⁶

Second, there is a coherent, legitimate normative question on the table.⁵⁷ As human beings, we have the capacity for reflection, the ability to step back from our mental states and to ask a wide range of questions, both theoretical and normative, about them. A person is able to ask, of any desire he happens to find in himself, whether it is motivated or basic, whether it is morally questionable, whether it gives him a weighty practical reason, whether it is prudent for him to work to strengthen or to weaken the desire’s influence on him, and so on. Along with

⁵⁶ For a discussion, see Arpaly and Schroeder (“Alienation and Externality”).

⁵⁷ Korsgaard complains that the “normative question,” the question what the ultimate source of normativity is, has “slipped through the cracks,” as philosophers discussed other, related questions, such as whether and why morality is rational, and what the real content of the moral law is. Korsgaard thinks that it is through confrontations with “problematic” desires, which seemingly ought to be ruled out, that we most often are driven to ask the questions that lead to this general and foundational normative question (Sources, 7–48).
these questions, a person is also able to ask—if the desire is basic—whether it is rationally (im)potent. Korsgaard puts things this way: the “reflecting self” is able to ask whether a particular “desire of the moment” gives him any reason to act whatsoever.

Frankfurt and Korsgaard clearly think this is a coherent and important question; both are emphatic that the answer to it is, in some instances, ‘no’. But even the subjectivist theorist who rejects the rational impotence thesis is able to think that this normative question is a coherent question, and it is hard to imagine why he would think that it is “illegitimate.” He is committed to giving the same answer in each case the question is raised: ‘yes, it does’. But that doesn’t make the question, in and of itself, a bad question. The second salient point of agreement is that each of the participants in the dialectic can agree that the question is coherent, though their answers are sometimes (importantly and tellingly) different.

Third, it’s necessary to make a qualification to the rational impotence thesis. Each participant should agree that the class of rationally impotent basic desires isn’t thoroughly empty. Even the most simplistic form of subjectivism grants that logically unsatisfiable basic desires are not able to generate practical reasons. What if an agent were to desire, in a basic way, to solve an irresolvable puzzle? The object desired is a logically impossible state of affairs. The right thing to say is that this basic desire is rationally impotent. Because nothing could count as a means to this desire’s satisfaction, it not only fails to generate a practical reason; it is incapable of ever doing so. The implication is that the real dispute is whether the class of logically satisfiable rationally impotent basic desires is thoroughly empty. Notice that the contested cases, the

---

58 Hubin raises this question and answers it in this way; he also provides the example (“Irrational Desires”).

59 What about contingently unsatisfiable basic desires? Are they rationally impotent? I don’t think so. They don’t generate any practical reasons, but they are not incapable of doing so. They would were an adequate means to arise.
purported examples of rational impotence mentioned so far, are logically satisfiable; not only could they be satisfied, logically speaking; they each do have adequate means. In Frankfurt’s example of a violent impulse, pulling the trigger will “do the trick.”

Disagreements

Significant disagreements arise immediately, and some of them cut to the very heart of long-standing disputes between competing schools of thought. At this point, I want only to give a taste of what’s at stake.

The first disagreement concerns whether norms for rational impotence exist. Consider Korsgaard’s claim that “Each impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action.” The subjectivist who rejects the rational impotence would not affirm this claim, at least not if basic desires are included among the “impulses” that must pass the relevant test. Though this subjectivist accepts that the normative question whether a basic desire is rationally potent is a coherent, legitimate question, he thinks the answer to it is always ‘yes, every logically satisfiable basic desire is rationally potent’. Consequently, it would be strained to attribute a “test of normativity” to him. (This so-called test would not be even minimally demanding, and so it would not be a test at all.) Frankfurt and Korsgaard, on the other hand, think that some basic desires are ruled out, which makes it natural to attribute to them a practical standard or a rational filter or a procedure of some kind, by which they are.\(^6\) (My aim, when arguing against my fellow subjectivist, will be to argue that such a “kind of test” does really exist.)

\(^6\) It is conceivable that the process by which rationally impotent basic desires are so recognized is not codifiable, as McDowell seems to think (“Virtue and Reason”). I don’t intend to discuss this possibility.
At the same time, Frankfurt disagrees with Korsgaard about the scope and the character of the relevant test, and the source of its normative authority. In Korsgaard’s theory, every motivational state, and so every basic desire, is subject to a proper test—a “correct procedure”—for the rational evaluation of desires. She does not exclude an agent’s deepest concerns, his loves and carings about, from this requirement. She accepts (what I take to be) a theory-defining Kantian thesis that a subjective motivational state is “external” or “alien” to the will. Given her commitment to the heteronomy thesis, Korsgaard thinks that the vantage point that you, as a reflective agent, are properly to adopt to evaluate an impulse is a perspective “over and above all your desires.”

Frankfurt disagrees. Frankfurt raises the standard “one thought too many” objection against Kant, and he would raise it against Korsgaard, too (Reasons of Love, 36-39). If you care about your own long-term welfare, you might step back from an impulse and ask, “Should I act upon this impulse? What does prudence say?” What prudence dictates counts as rational advice for you. In this sense, at least, it is proper for you to reason from and not to this basic desire: you can treat it as an ultimate source of practical reasons. Frankfurt also denies the spirit of the theory-defining Kantian thesis in the following quotation:

More evidence: Korsgaard claims that it is “better” for an agent to love his values (Sources, 103); this way there is a correspondence between the principles of choice that make up her will and her “heart.” I don’t think a “value” is a legitimate source of practical advice unless it is linked to the agent’s desires, and an agent’s cares and loves provide her with her most significant normative guidelines.

For an application of this objection to Korsgaard, see Gibbard (“Morality as Consistency”).

Frankfurt tends to identify an agent’s loves—the things he cares about, inescapably—as a necessary ground of practical reasoning. I disagree with Frankfurt about this point. I think caring about is sufficient for rational agency and personal autonomy, and also that it is conceivable that an agent’s caring abouts are more significant ruling passions than his loves. For a similar view, see Michael Bratman’s reply to Frankfurt’s Tanner Lectures (“A Thoughtful and Reasonable Stability”).
the most basic and essential question for a person to raise concerning the conduct of his life cannot be the normative question of how he should live. That question can sensibly be asked only on the basis of a prior answer to the factual question of what he actually does care about. If he cares about nothing, he cannot even begin to inquire methodically into how he should live; for his caring about nothing entails that there is nothing that can count with him as a reason in favor of living in one way rather than another (Reasons of Love, 26).

In summary, Frankfurt thinks that the practical vantage point from which ‘you’ ought to evaluate the rational potency of basic desires is not independent of all of your basic desires; it is defined in terms of some of them. The “desire of the moment” is to be evaluated from a perspective that centrally consists of what you care about.64 This is generally a substantive, not a thoroughly formal, test.

The standard objections could start to fly back and forth between Kant and Korsgaard, on the one hand, and Frankfurt and me, on the other. I want to hold off most of this until later chapters, to get clear about what is at issue. For all the disagreements, Kant, I submit, is a natural dialectical partner for Frankfurt. For the time being, I’ll mention two reasons why. Frankfurt’s theory is an attempt to capture a host of intuitions about practical reasoning and personal autonomy. Frankfurt wants to make sense of the idea, of central importance in Kant’s thinking, that a person is not merely “an arena in which desires struggle for satisfaction,” but has the capacity for active self-governance.65 Frankfurt also agrees with Kant that we often confront, in

64 I don’t mean to deny that an agent’s practical point of view will include mental states other than desires. Intentions will be formed and personal policies adopted in the pursuit of what the agent cares about. Various beliefs might explain why the agent has chosen to form these intentions and adopt these policies, and so—it seems to me—also represent “the agent.”

65 To account for various intuitions about free action, Kant appeals to a robust theory of the will, whereas Frankfurt bootstraps. One intuition Frankfurt bootstraps to is that an agent is not merely an “arena in which desires struggle for satisfaction,” but plays an active role, in at least some instances, in bringing action about. How to make sense of this idea on subjectivist grounds? Velleman argues that the “standard” story in the philosophy of action for what happens when an agent acts is that there is something the agent wants and an action that the agent believes conducive to its attainment. His desired end, along with his belief in
the course of reflection upon what to do and how to live, various ‘musts’—actions, or whatever, that are not only “best” but in some sense “necessary.” Some dictates of the will are “counsels”—they are good advice, which we ought to consider taking, but which do not compel obedience: “Prudence tells me to give up red meat altogether. But I’ll let myself indulge from time to time.” But we also, both Kant and Frankfurt think, come up against dictates that we cannot set aside—“commands.” Similar to the compelling necessity of the rules of logic for theoretical reasoning, we confront “volitional necessities” (“Getting Things Right,” 30-32).

Frankfurt’s aim is to capture both of these intuitions—and maybe more—without going Kantian in any significant sense. These are some of the intuitions he aims to bootstrap to, and

---

66 The terms “counsels” and “commands” are Korsgaard’s.

67 I suspect that the reason that Frankfurt has not made more progress with respect to the rational impotence thesis is because he has tended to focus on the concept of love, which is, in his theory, a species of caring about: it is caring about, inescapably. Loving creates a different type of ‘must’ than a norm of rational impotence: if you can’t help caring about someone, then you can’t help seeing that person’s welfare as a thing to value. The voice of the desire for her welfare is a voice you can’t, in a psychological sense, get rid of.
his commitment to this intuitions help to create the tension in Frankfurt’s account. Kant, then, is a natural dialectic partner for Frankfurt. It’s beyond the scope of my project to discuss each type of “volitional necessity.” I will discuss the ‘must’ implicit in the claim “I can’t rightly treat such and such a desire as reason-giving.”

Korsgaard’s theory and argument, it turns out, are of special relevance to my proposal. I appeal to the concept of a personal ideal to refine Humean subjectivism, she to refine Kant. Who gets the contested ground? My final chapter will address this question. I’ll argue that she appeals to the correct concept, or at least something close to it, but gives—because of her commitment to the heteronomy thesis—valid personal ideals the wrong normative grounding.

In the following two sections, I will outline an initial defense of the two premises in the anti-subjectivist challenge, with an eye to clarifying the debate and getting initial considerations onto the table.

The Case for the Rational Impotence Thesis

I will describe two, sketchy arguments for the rational impotence thesis. The first is largely an appeal to purported examples of rational impotence; the second is an argument with a distinctly Kantian flavor. It claims that a reflective agent that does not subject all of his basic motivational states to a proper test of normativity—a test that, at least in principle, rules out some basic desires—is not a fully reflective or fully autonomous agent. As I see it, the second argument goes further than the first. The first is neutral between different forms of anti-
subjectivism, both realists and Kantians are able to accept it; the second commits itself to theory-defining Kantian claims about rational agency and personal autonomy.

The Appeal to Purported Examples

The first argument appeals to an experience a reflective agent might have in the context of practical deliberation. It goes something like this. As Cohon remarks, it seems coherent to ask, of any particular desire, whether it generates a (pro tanto) practical reason (“Roots,” 63). It is, someone might say, an “open” question. And it would be striking, the argument suggests, if there were never a basic desire for which the answer is ‘no’. Moreover, don’t we already recognize, pre-theoretically, that some basic desires are rationally impotent? Cohon, again: “Some desires we classify as reasons to pursue the objects of desire; some we discount altogether” (63). From an appeal to plausible examples, we can make clear that there are rationally impotent basic desires. (More about the plausible examples in a moment.) Why not conclude, the argument might say, that what seems to be the case really is? Why distrust our intuitions about what counts, and what doesn’t, as a legitimate source of a (pro tanto) practical reason and our common way of classifying desires? The argument concludes that we shouldn’t distrust these intuitions: the purported examples are legitimate examples; there are, in the very least, some rationally impotent basic desires.

What are the plausible examples? I suppose it is difficult to anticipate which examples readers will find most plausible. Different philosophers have suggested different examples. Let’s take the scattershot approach and suggest a handful of examples, with an eye to choosing some that are interestingly different from each other. At least one of these might strike the reader as a
highly plausible, or even legitimate, example; and if there is even a single rationally impotent basic desire, this single example is sufficient to entail the rational impotence thesis.

Cohon puts forward that neither her strange whim to stick her finger in a gooey substance nor her “hateful” desire to stick it to an unpleasant colleague give her a positive reason to stick anything anywhere. The two cases are different, she thinks. There is simply nothing going for the activity of sticking your finger in a gooey substance; the object of that whim, she seems to think, is disgusting and has no value whatsoever and so generates no practical reason whatsoever. And, as for the hateful desire, there is a positive reason against sticking it to another person: it’s immoral. So the one desire is rationally impotent because it has nothing going for it; the other because it has something (sufficiently serious) going against it.

I’ve already mentioned several other examples: Harry Frankfurt’s examples of the unwilling addict’s present desire to take drugs and his version of the parent with a cherished child and a lethal desire. The second example is, perhaps, inspired by an older one: Gary Watson’s example of a loving mother with a violent impulse to drown her beloved infant son (“Free Agency”). Watson also provides the colorful example of a person who wants to bash his squash partner in the head with a racket after losing to him.

Here are a few more candidate for rational impotence. A decent, thoughtful person feels a rebellious desire to make a loud and disruptive noise in the middle of a solemn ceremony—a funeral, let’s say. A happy person who feels a strange urge to veer into oncoming traffic might reject the suggestion of the urge: “I’m happy; I have absolutely no reason to kill myself.” Also, a pacifist who suffers an unprovoked and aggressive shove in the middle of a pick up basketball game might feel, naturally, a vengeful impulse to shove back. Regardless of what he actually does, he might think his vengeful desire does not give him a practical reason to strike back. If he
does strike back, the desire for vengeance would explain his behavior, he might think, but it would not justify it in any way, not even minimally.69

I take it that the author of the argument from experience would like the reader to identify with one or other of the possible examples: “Haven’t you, in the course of your life, come across desires such as these, which you were unwilling to give a voice to in your deliberation?” Hence, this is the argument from the experience of practical deliberation. Once you identify with one or other of the examples, the author thinks you will then draw the conclusion that our best theory of practical reasons must accommodate the legitimate examples and their logical implication, the rational impotence thesis.

These examples are more suggestive than well-developed. But in all fairness, the appeal to purported examples must begin somewhere, and it strikes me that the examples provided are sufficient to begin discussion. (If you don’t think I’ve given the most compelling candidates, fill in your favorite.) The question is whether any of these purported examples (or your favorite one) is a legitimate example of rational impotence. I’ll discuss them at greater length in chapter 3.

I’ve already said it: I accept the rational impotence thesis. To critics, I would submit that, in the very least, this sketchy argument does have something going for it, at an intuitive level. People do experience some very strange, and some very horrific, desires. But I don’t think an appeal to purported examples is sufficient, by itself, to establish the rational impotence thesis. Strange and distressing desires raise interpretative questions; they call out for theoretical understanding. Frankfurt stipulates that his desire to shoot his son is “really a desire,” but of course it is conceivable that we often misinterpret our mental states or make mistaken judgments.

---

69 If we isolate the pacifist’s instrumental mechanism, we could say it performed as it should: it attached means to end. There wasn’t any mistake made in the performance of this mechanism. But, the pacifist seems to think, it operated on behalf of a desire that had no right to put it to use.
about them. I’ll discuss misinterpretations and mistaken judgments in chapter 3. I find it very likely, antecedently, that many of the purported examples of rational impotence will admit of several interpretations. It’s not easy to make sense of what’s really going on in our heads. But I’m disinclined to interpret them all as something other than what Frankfurt is gesturing at.

More importantly, I would like to highlight that the purported examples are interestingly different. Think of Frankfurt’s own examples. His violent urge, we are to believe, is a bizarre interloper. Frankfurt describes himself as both distressed by its content and completely in a fog about its etiology: “No doubt it signals God knows what unconscious fantasy” (“Taking Ourselves Seriously,” 12). Frankfurt describes some desires as “hot surges of anarchic emotion,” “mental tics, twitches, and convulsions,” and “psychic analogues of seizures.” (8-9). But these descriptions don’t fit his canonical example. The unwilling addict’s present desire for heroin is yet another dispatch from his old nemesis, his addiction, which has defeated him, he knows, time and time again. It isn’t “out of the blue.” He knows well enough where it came from. Even if he doesn’t know the full physiological and chemical story about addiction, he recognizes part of the etiology: his choices and subsequent behavior got him addicted. Another difference is that the addict does recognize that acting upon the present desire for heroin has something going for it: the immediate high or, if that no longer comes, the alleviation of his distress. There doesn’t seem to be any recognizable good that would come from pulling the trigger.

I draw a tentative conclusion. An appeal to something simple, such as a desire’s fleeting character, is not sufficient to secure the rational impotence thesis. Some strong candidates for rational impotence are complex and persistent desires, and some do have some allure. I think an

---

70 This means that the addict could conceivably have a practical reason to satisfy the addictive impulse as a means to another desired end: pleasure-acquisition or pain-avoidance. It doesn’t follow from this that the addiction itself is rationally potent.
unwilling addict might well get caught up in the allure of his addictive impulse, even though he is committed to a personal ideal that forbids him to. What holds Frankfurt’s two cases together, in my view, might be nothing more than that the agents are each rejecting a desire as a source of a practical reason—that and, if these are legitimate cases, a valid norm of rational impotence.

In any case, in the third chapter, I’ll discuss the case to interpret these examples in a way that does not support the rational impotence thesis.

Intuitions about Autonomy and Reflective Agency

A second argument for the rational impotence thesis appeals to more robust intuitions about personal autonomy. Korsgaard claims that “the reflective mind cannot settle for . . . desire, not just as such. It needs a reason . . . our impulses must be able to withstand reflective scrutiny” (Sources, 93). If this remark were intended as a bald objection to subjectivism, it would be question-begging: it simply asserts that basic desires are not rationally potent.

But let’s not interpret it that way. A thesis such as the claim that any motivational state—and so any basic motivational state—is external to the will tends to cut to the heart of questions about the nature and purpose of practical reasoning. This kind of Kantian claim, which clearly shapes much of his, and much of Korsgaard’s theorizing, should be judged, in part, on how the theory built around it does. And, in fairness, Kant and Korsgaard supply sustained attempts at articulating these standards. Korsgaard appeals to intuitions about personhood and autonomy to argue for this claim. For the time being, let’s rely upon something less than a full-blown Kantian theory. Let’s start with some intuitions.

Begin by thinking of Willy. Willy is self-reflective enough to be aware of his various basic desires, and he is adept at satisfying them. Willy rarely falls afoul of the standard principles
of decision theory: he generally has a determinate preference ranking, and knows to plunk for one when that is the thing to do. And once Willy has settled on what ends he is to pursue, he follows the various accepted principles for the weighing of practical reasons against each other: the principle of transitivity, invariance, dominance, and so on. He is a decision-theorist’s dream—though Willy is not a “pure maximizer”: to the degree that Willy cares about various thick, action-guiding ideals, he employs a language that makes qualitative distinctions that often shape what he thinks it best, all things considered, to do. He counts some behavior as ignoble or lewd or otherwise good and bad, and these behavioral norms function deontically (perhaps as side constraints) for him, often bringing him to rule out various actions or action-types as practical options for him. In short, Willy is not a thorough-going “wanton,” a person who does not participate in the processes leading all the way to action. He’s involved, and he’s a pretty sophisticated practical reasoner.

But there is something Willy doesn’t do. Willy does not call his basic desires into question, at least not in the sense of subjecting them to a test for rational potency. He recognizes that one basic desire might make it more difficult, instrumentally speaking, to satisfy another, for instance; but this is not, as the criticism goes, deep reflective criticism. Willy treats each of his de facto basic desires as input into his practical deliberation—without a thought, as it were. He ignores the voice of a basic desire only if and when doing so is the most effective means to something he wants even more.

The argument proceeds by remarking that Willy’s life and activity, though not utterly wanton, displays an uncritical edge. Willy is a creature capable of reflection, and he exercises that reflection, and well—but only up to a point. He doesn’t, as Korsgaard puts it, “press reflection as far as it will go.” He does not ask any (non-instrumental) normative questions about
his basic desires. The normative question ‘I do want this thing for its own sake, but do I have any reason to acquire it?’ is a coherent question, which arises in the context of reflective agency, which has—potentially—some very serious practical ramifications. But Willy doesn’t ask it. As we’ll see, Korsgaard submits that he ought not only to ask it, he needs an answer to it.

The Kantian might also remark that Willy’s life suffers from a degree of heteronomy. What sets human beings apart from the other objects in the world, even complex animals, is our ability to ask normative questions. This is part of what makes a human being a person. To stop short of asking normative questions is to ignore the call of your very humanity, or your very personhood. The process of maturation from a small child to a fully autonomous agent is a matter of reflecting upon more and more, scrutinizing even more and even deeper assumptions, motivations, and patterns of reasoning. To attain full autonomy, reflection upon one’s basic desires is necessary. A fully autonomous agent would step back from each of his motivational states and call into question each of his basic desires, to see if they really give him reasons for acting. In short, each of our basic desires generates nothing more than a prima facie practical reason, which has to stand before the judgment seat of our wills. And to deny the rational impotence thesis is tantamount to the denial of a central aspect of our best conception of human agency.

71 It is notable that this is also a central theme in Frankfurt’s own work, and has been since his seminal article, “Freedom of the Will and the Concept of a Person” in 1971.

72 T.M. Scanlon puts the same point in this way: a desire is a “seeming”—to desire something is for it to seem as though you have a practical reason to seek it—but a desire is not yet an “assessment”—a desire that has passed a test of reflection—let alone an “opting”—a judgment that it is the thing to pursue, all things considered (“Reasons and Passions,” 174-179).

73 Is a third argument lurking? Given the vagaries of human desire, to give over even a small degree of rational potency to such things as whims or urges is dangerous; it is to give over normative status to what has already proven to be an unruly and fickle motivational state that does not proportion its intensity or its
In one sense, these Kantian arguments are old news. If dropped into the middle of this debate, the heteronomy thesis seems question-begging, and the idea that the subjectivist theory is forced to admit that an agent might have a practical reason to do a nasty or weird thing isn’t the least bit novel. James Dreier captures the standard mood among subjectivists, the possibility that it could be rational for an agent to behave in ways we find morally objectionable or simply ludicrous is “a disappointment we’ll have to learn to live with” (“Humean Doubts, 33). I think the strength of my own proposal, which appeals to personal ideals and the substantive practical advice they issue, will come through when compared to Korsgaard’s Kantian proposal.

The Incompatibility Thesis

Is the incompatibility thesis true? Is a clear-headed, consistent subjectivist compelled to reject the rational impotence thesis? I don’t think so, and I should say why.

Whether the incompatibility thesis is true depends upon how subjectivism should be defined. What is the subjectivist theory? What are its defining characteristics, and to what is it inescapably committed? Someone could, of course, propose to define subjectivism in such a way that it will then become rational impotence. How, then, does a subjectivist guarantee that it will not come to have the status of ‘the desire to act upon’? Unless the subjectivist indicates what grants (greater) rational potency to a basic desire, we could worry that a very unseemly one might come to have the relevant property, or set of properties. For instance, what if rational potency were in direct proportion to intensity or causal power? We could imagine that a present urge to take heroin is very intense and exerts significant causal power. It would become a ruling passion. If a theorist were to admit that it is rationally potent, wouldn’t that open the door to its becoming a ruling passion? I don’t think so. I reject the idea that rational potency is in direct proportion to intensity or causal power. Frankfurt’s account of ruling passion does, too. I do, though, think there is a truth, a relatively small truth, in this objection. In my subjectivist theory, Frankfurt’s violent impulse could come to have rational potency: if he comes no longer to care about being a decent father or any other personal ideal with a norm of rational impotence that rules it out. It could also become a ruling passion, if Frankfurt were to undergo an exceedingly large change in desiderative profile.

Also, Fehige thinks the case for subjectivism is very strong, and doesn’t yet see a compelling reason to adopt a competing theory, though he admits that perhaps there are compelling reasons that he hasn’t yet noticed: “Far be it from us to jump to conclusions” (“Instrumentalism,” 65-66).
that it accords rational potency to every basic desire. If we were to accept this definition, we would, of course, have to accept the incompatibility thesis, too. Simply stated, I do not accept this definition. What follows is best construed as an argument not to: the most plausible theoretical commitments of the subjectivist tradition do not imply the incompatibility thesis. Subjectivism is committed, I think, to the claims that basic desires are the only source of practical reasons and that there aren’t any “extra-subjectivist” constraints upon practical reasoning, in a sense I will specify.

This conclusion gets me only to the claim that the rational impotence thesis is logically consistent with subjectivism. I will go on to argue, at the end of the third chapter, that the rational impotence thesis is not only consistent with the best aspects of subjectivist thinking, but that subjectivists have positive reason to accept it.

The Core of Subjectivism

Let’s begin with the two basic tenets that make up, to my mind, the heart of subjectivism. The first is the claim—which I take to be “theory-defining”—that practical reasons—all of them—are derived, ultimately, from among an agent’s own subjective, contingent basic motivational states. This claim has been called “the desire-based reasons thesis.”

The second basic element of subjectivism is an instrumental principle. An instrumental principle transmits normativity from ends to means. It states, as a first approximation, that if you have reason to bring about an end, then you have reason to take the adequate means. Any

---

55 See Darwall (Impartial Reason), and Hubin (“What’s So Special?”).

56 In chapter 3, I will consider a minor qualification to this description of the instrumental principle.
plausible theory of practical rationality will include an instrumental principle; means-end reasoning is an indispensable aspect of practical rationality. That said, the instrumental principle might be described differently from theory to theory, in ways that reflect substantive differences between them. Within a subjectivist theory, which grounds normativity in “elements of an agent’s S,” the instrumental principle communicates normativity from rationally potent basic desires to the adequate means to their satisfaction.

Take what I’ve said so far. Does it imply that the incompatibility thesis is true? No. The basic tenets of subjectivism do not entail the incompatibility thesis; that is, they do not entail that the rational impotence thesis is false. Let me say a bit more about the desire-based reasons thesis. It has two parts. First, it states that practical reasons don’t derive, ultimately, from any source other than a basic motivational state. Second, given that subjectivism is not a form of skepticism about practical reasons—it does think that some practical reasons exist—the thesis also implies that there are rationally potent basic desires. In summary, practical reasons derive only from desires, and some desires do generate practical reasons. This does not imply, in and of itself, that every basic desire generates a practical reason.

So, there needs to be an argument for the incompatibility thesis; it doesn’t follow with necessity from the basic constituents of the subjectivist theory. This doesn’t settle the matter, though. The two basic tenets of subjectivism don’t make for a fully-fledged theory of practical rationality, and it could be that in the process of theory-construction from the basic elements to a fully-fledged theory that the truth of the incompatibility thesis emerges.

Why think that the two basic elements don’t make for a fully-fledged theory? The primary purpose of a theory of practical rationality is to provide rational advice. Say I desire, in a

---

77 For a discussion, see Hubin (“What’s So Special?”), and Dreier (“Humean Doubts”).
basic way, to go to my son’s soccer game and to go to my sister’s wedding, but I can’t do both.

The basic elements of the subjectivist theory do not have the theoretical resources to give me any practical advice about how to resolve the conflict between the conflicting, *pro tanto* normative considerations that derive from my competing basic desires. The basic elements of the theory say only that I have a practical reason to do one action and a practical reason to do the other; they do not bridge the divide between the practical reasons that derive from an agent’s rationally potent basic desires and the final ‘ought’ of rational advice.

The problem isn’t simply that the basic elements do not have the resources to declare a winner; it’s conceivable that, in some cases, there isn’t a single course of action that is the best: several weightiest practical reasons could be equally weighty, all things considered. The problem is that the basic elements of the subjectivist theory don’t even indicate whether going to the pitch or to the church are equally advisable or whether rationality calls for my simply plunking for the game or for the wedding; the basic elements simply have nothing to say.

To generate a fully-fledged theory, the basic tenets need to be supplemented—by weighing principles, for instance. There have been, to say the least, a myriad of proposals for how to get, properly, from practical reasons to the final ‘ought’ of rational advisability. Frankfurt’s appeal to the concept of caring about is one proposed supplement. He accepts a general thesis that an agent’s ruling passions generate practical reasons that often outweigh the practical reasons derived from “mere” desires.

The possibility exists that whatever is added to the core of the theory will imply—by itself, or in combination with the basic elements—that the rational impotence thesis is false. This mere possibility isn’t itself a reason to accept the incompatibility thesis, of course. It isn’t, so far
as I can see, even a reason to suspect that the incompatibility thesis is true. But it does indicate that my work isn’t done. I need to track the subjectivist theory’s theoretical commitments further.

This isn’t the place to discuss all of the ways that subjectivists have supplemented the basic tenets of subjectivism. My strategy will be more circumscribed. I will discuss several claims that are characteristic of subjectivist theorizing and that might seem, initially, as evidence for the incompatibility thesis.

Common Subjectivist Ideas

I take it that there is some grist for the defender of the incompatibility thesis. Subjectivists do say some things that seem, at least initially, to cut off the very possibility of accepting the rational impotence thesis. I will discuss two claims that are characteristic of subjectivism and that could be taken as evidence that the theory cannot countenance rationally impotent basic desires. Subjectivists deny the existence of categorical practical reasons and categorical norms of practical reasoning—and so they tend to castigate “categoricity” in practical reasoning. Also, they often construe their own theory as the “instrumental theory of practical rationality.” These are stock themes in subjectivism, themes that Frankfurt seems to fly in the face of when he claims to categorically reject his violent impulse on non-instrumental grounds.

Also, there simply isn’t much precedence within the literature for the view Frankfurt and I accept: a subjectivist theory that accepts the rational impotence thesis.78 “Can’t we learn something about a school of thought by noticing what it doesn’t say?”79 It’s not as though

---

78 Hubin talks of a “value screen” that basic desires are passed through. But Hubin does not intend, he tells us, to indicate what an agent’s values, in the relevant sense of “values,” are (“Desires, Aims, and Values”).

79 Hubin wonders whether subjectivists have been more focused on arguing against competitors than on refining their own theory.
subjectivists have lacked opportunity to endorse the disenfranchisement of basic desires. On the contrary, when subjectivists are presented with cases in which an agent experiences a bizarre or despicable desire, they generally treat it—if it is stipulated to be a basic desire—as rationally potent, no matter how bizarre or despicable. In this, subjectivists take their cue from Hume who famously says, in *The Treatise of Human Nature*, “‘tis not contrary to reason to prefer the destruction of the whole world to the scratching of my own finger.”

A sharp-tongued critic might remark, “How could a theory that does not judge Hume’s crazy preference irrational possibly claim that some basic desires are rationally impotent?”

Subjectivists do standardly interpret the desire-based reasons thesis inclusively. The referent of ‘desire’ is a broad category of motivational states. On the linguistic face of it, there seem to be a wide variety of states that we commonly distinguish from “desires.” On the one hand, some conative states—cares, loves, enthusiasms, attachments, aspirations, insistent yearnings, thoughts about what one ought to do—seem to be more stable, complex, or important to us than “mere desires.” On the other hand, urges, whims, and cravings seem to be more fleeting, simple, or trivial than our “average” desires. The subjectivist theory categorizes any of these ‘goal-directed’ motivational states as “desires” in the relevant sense. What the desire-based reasons thesis is intended to articulate is that an agent’s practical reasons, all of them,

---

80 Subjectivists commonly assimilate these cases to what Herman calls the “plural interest model.” The basic desire is treated as rationally potent, and the practical reason it generates is compared to other practical reasons the agent is imagined to have (“Bootstrapping”).

81 See Fumerton (*Reasons and Morality*, 129-130).

82 I suspect that people often call a motivational state an ‘urge’ or a ‘whim’ to indicate that it is not something that he really cares about. The descriptions reflects, that is, a judgment about the motivational state and its fit with the agent’s practical point of view.
ultimately derive from the agent’s own subjective, contingent, and actual motivational states.\textsuperscript{83} A consideration can be a practical reason for you, the theory claims, only if it is properly connected to one of your own basic desires—wherever along the spectrum, from simple and fleeting “whims” to complex aspirations and abiding loves, this basic desire happens to fall.\textsuperscript{84} The relevance to my discussion is that this inclusive tendency gives the impression that the subjectivist theory can’t and won’t rule anything out.\textsuperscript{85}

I agree, by the way, that it would not be consistent and clear-headed for a subjectivist to rule out a whim or a craving on the mere grounds that it is bizarre or fleeting. (My proposal doesn’t.) I think it is easy to think of bizarre and fleeting basic desires that are rationally potent. Recall my defense of the fleeting urge to stick a finger in goo. It is conceivable that an agent might have a personal ideal that prescribes sober seriousness, so that the agent who cares about it is properly unwilling, on my account, to treat a silly whim as reason-giving. In my view, it would take a story such as this for the whim to be rationally impotent for the agent. The whim isn’t ruled out on the mere grounds that it is trivial, fleeting, or even strange.

\textsuperscript{83} See Hubin (“Hypothetical Motivation”).

\textsuperscript{84} What if a subjectivist were to attempt to argue against the rational impotence thesis on the grounds that, though urges and whims are rationally impotent, they are not desires? A desire is a motivational state that meets a “desirability criterion,” and they do not. I do not intend to construe this as a subjectivist theory. In my discussion, the term ‘desire’ is being employed in a highly inclusive sense. It includes psychological states that we generally call urges, whims, pro-attitudes, valuations, and caring abouts. What these various psychological states share, however they differ, is goal-directedness: a particular world-to-mind direction of fit, which fits them for producing action (in conjunction with appropriate beliefs).

\textsuperscript{85} For Frankfurt’s discussion of Hume’s famous remark about the destruction of the world and his own finger, see “Getting Things Right” (28-34). Frankfurt seems critical of Hume, initially; but his final judgment is not that the bizarre preference is intrinsically irrational, but that you and I, given what we find important, can’t “identify” with it. See Hubin’s discussion of unintelligible desires (“Irrational Desires”).
The inclusive tendency of subjectivism does place a theoretical burden on a theorist, such as myself, who is a self-described subjectivist who accepts the rational impotence thesis. But it doesn’t get to the bottom of things. In a moment, I’ll discuss why subjectivists are so inclusive.

Unsound Arguments for the Incompatibility Thesis

Before I move on to the more plausible arguments, let me rule out some unsound arguments for the incompatibility thesis. There is the possibility that the anti-subjectivist challenge could be grounded in a misconstrual of subjectivism. The anti-subjectivist challenge would lose much of its force if we were to discover that it is aimed at implausible forms of subjectivism that few subjectivists actually accept.

For instance, imagine a subjectivism—reductivist in spirit—that treats rational potency as either grounded in or directly correlated with phenomenological intensity: the more intense the desire is felt, the greater its normative authority. If objections to subjectivism were directed at this version of subjectivism, they would miss their target. A subjectivist doesn’t have to—and most contemporary subjectivists don’t—accept this correlation. Many subjectivists think that a weakly felt basic desire might have significant normative authority: the unwilling addict’s desire not to take drugs might be a very “calm passion,” or derive from one, and yet have greater normative authority than even a very intense addictive urge. Though speaking softly, the unwilling addict’s desire not to take drugs carries a big, normative stick.  

Also, we could imagine another reductivist subjectivism that grounds a basic desire’s degree of authority in its causal efficacy. This, too, is not a plausible form of subjectivism.

86 Given that not every desire is felt, and not every desire leads an agent all the way to action, these two reductivist claims are both consistent with the rational impotence thesis. I don’t intend to challenge the incompatibility thesis on these grounds, though, since I think these reductivist claims are both false.
Frankfurt, among others, is committed to thinking that a basic desire that generates a final ‘ought’ of rational advice might lose out in the struggle to bring the agent all the way to action; the unwilling addict’s desire not to take drugs ought to win out in this struggle, even if it doesn’t.

Frankfurt’s appeal to the concept of caring about is intended, in part, to capture these intuitions about rational potency. The addict “merely wants” the heroin, whereas he “cares about” not taking it, and—other things being equal—caring abouts generate weightier practical reasons, independent of their phenomenology and causal power.

Here’s a second bad argument for the incompatibility thesis. Philosophers sometimes ask whether a basic desire “is” a practical reason. Korsgaard, for instance, raises this question, and gives an answer to it; she remarks that “a desire isn’t a reason.” I don’t know whether she means to attribute the thesis that a basic desire is equivalent to a practical reason to the subjectivist theory, but let’s imagine that someone else does. If this “equivalency thesis” were true, then the incompatibility thesis would be true: the moment a basic desire came into existence, so, too, would a practical reason, since they are the same thing. The rational impotence thesis would be incompatible with subjectivism. But a subjectivist isn’t compelled to accept this equivalency. First, a desire is a mental state, whereas a practical reason is a normative entity. Second, a subjectivist could adopt, and many implicitly do adopt, the thesis that a rationally potent basic desire is the “root” or “ground” or “ultimate source” of practical reasons. A subjectivist, even if he does not accept the rational impotence thesis, is able to adopt the thesis that a basic desire “generates” or “creates” practical reasons, but is not equivalent to them.

---

87 Also see Scanlon (What We Owe Each Other, 51).
It is also important not to misconstrue subjectivism by identifying it too closely with what Hume himself thinks. Though contemporary subjectivism clearly draws inspiration from Hume, it departs from him in at least two significant ways.\(^{88}\) (Contemporary subjectivism is often called “neo-Humeanism” to reflect this.) First, subjectivists generally abandon Hume’s theory of motivational states. Hume conceives of motivational states, “passions,” as feelings; contemporary subjectivists generally accept the more plausible thesis that a person is able to have a motivational state, one relevant to rational advice and the rational appraisal of action, without feeling it. (You ask: “How, then, do we know of such states?” We might discover the existence of a basic desire not only phenomenologically, but from its causal influence upon behavior or from its production of emotions that we do feel.)

A second deviation from Hume concerns Hume’s reduction of reason to theoretical rationality. Hume thinks of reason as aiming at the truth, and so criticisms of irrationality are only ever criticisms of theoretical irrationality. According to Hume, an agent’s behavior can be irrational, but only if the desire that causes the behavior is a motivated desire that has been produced on the basis of unwarranted (means-end) beliefs. Hence, his theory implies that no one can be guilty of what has been called ‘true practical irrationality’.\(^{89}\) Say that an agent is motivated, in a basic way, to achieve an end, believes rationally that a particular action is the best means to it, does not think that there is any overriding countervailing practical reason, but fails to perform this action. This agent is guilty of a form of non-theoretical irrationality: “means-end irrationality.” A theory of practical rationality should make conceptual room for this possibility:

\(^{88}\) For helpful discussions, see Millgram (“Was Hume a Humean?”); as well as Smith (The Moral Problem, chapter 4) and (“The Humean Theory of Motivation”).

\(^{89}\) The phrase is Korsgaard’s (“Skepticism about Practical Reason”).
were an agent to behave this way, he would be guilty of practical irrationality. Hume’s theory does not. But contemporary subjectivists not only accept that means-end irrationality is a form of practical irrationality, it generally serves as a paradigm example in their theorizing.

The idea that subjectivism is a simple, even simplistic, theory has been encouraged, I think, by the way subjectivists themselves talk. Even among subjectivists, one commonly hears references to the “standard” model of agency, which tends to be very minimalist, operating with only a minimum of basic concepts. The sparsest includes only belief, desire, and action. This model is often a good heuristic. If you would like to distinguish between various forms of valuation, or to include the concept of “norm-acceptance” or “intentions, plans, and the decisions from which they normally arise” or “identification with a desire” at the basic level, fundamental to any adequate philosophy of action, then it makes sense to indicate how these mental states differ from garden-variety beliefs and desires. This is good, philosophical practice: comparing what we don’t fully understand to what we do. The common use of this “standard model,” though, might give the false impression that subjectivism is nothing more than the minimalist theory.

Let’s turn, then, to more legitimate grounds for the incompatibility thesis.

---

90 Watson distinguishes between desires within an agent’s valuational and motivational systems (“Free Agency”). The next three references are to Gibbard (Wise Choices, Apt Feelings); Bratman (Intentions, Plans, and Practical Reason); and Frankfurt (Reasons of Love).

91 Hume is a skeptic, of a sort, about the will: for him, it is simply the subjective impression of the working of other parts as they yield up action (Treatise, 399). Frankfurt follows Hume in his earliest theorizing, when he identifies the will with “effective desire” (“Freedom of the Will and the Concept of a Person”). But Frankfurt deviates from this practice. More pointedly, Frankfurt is particularly concerned to account for the distinction, in our mental lives, between what counts as active behavior and passive experience.
Intrinsic Rational Criticism of Basic Desire

It helps to explain why subjectivists treat a wide range of motivational states as rationally potent, even Hume’s bizarre preference. Subjectivists accept Hume’s characterization of basic desires as “original existences” and claim that, as such, they do not admit of (valid) intrinsic rational criticism. This, I take it, is theory-defining.

The thesis that a basic desire is an “original existence” is commonly interpreted in terms of the ‘world-to-mind’ fit of desires. Consider the distinction between beliefs and desires. Beliefs have a world-to-mind direction of fit. Things go well for a belief, so to speak, if the content of the belief conforms to the way the world is. Desires, on the other hand, are goal-directed states. Things go well for a desire, so to speak, if the world conforms to them. The role that a desire plays in bringing about action is to press us—to push or pull us—to bring about states of affairs: to motivate; whereas the role played by belief is to indicate what needs to be done to bring about these affairs.

Now consider the case against a false belief. A false belief does bear a defect, so to speak: the propositional content of a false belief fails to conform to the way things really are. Belief aims at the true. A false belief, therefore, is not living up to its calling. It is “supposed” to correspond to a reality that is independent of it; but it doesn’t. Anyone who makes these remarks is, of course, committing herself to the existence of “truth-makers”—with respect to empirical beliefs, to the existence of the natural world, a mind-independent reality to which her empirical beliefs might fail to conform. Now consider a basic desire. In what way, subjectivists generally

---

92 This way of describing motivational states has been challenged: Searle (Intentionality, 8, n.), Millgram (“Current State,” 5), and Fehige (“Instrumentalism,” 56). Searle argues that ordinary objects don’t have an intrinsic direction of fit, and there isn’t any reason to think that mental states differ from ordinary objects in this way. Fehige says that the description of a desire as “what the world should track” appeals to the word ‘should’, which doesn’t admit of a trouble-free interpretation.
wonder aloud, could a basic desire reveal any kind of defect analogous to the falsity of belief? One answer, a realist answer, is that a basic desire “aims at the good” and, if its object is not good, it fails to conform to normative reality. But a subjectivist is committed to denying these claims; both that basic desires, generally, aim at the good, and that there exists a normative reality—a set of “truth-makers”—that their objects might or might not conform to. This is part of what his theoretical modesty amounts to. In the very least, rejecting the existence of a normative reality, a realm of normative facts, to which a desire is to conform isn’t akin, epistemically, to rejecting the existence of garden-variety descriptive facts.

A chain of reasoning might lead from this distinction between beliefs and desires to a rejection of the rational impotence thesis. A subjectivist might appeal to another disanalogy between believing and desiring to motivate his simpler theory. Having conflicting beliefs seems to press you to get rid of one of them; having conflicting desires, by itself, does not. To be of one mind, with respect to beliefs, requires consistency in believing. If you have conflicting beliefs, the only route to being of one mind is to reject one or both of the inconsistent beliefs. Theoretical rationality presses us in the direction of a consistent belief set. Practical rationality, arguably, requires a type of unity, too. A paradigm example of practical irrationality is a person who pursues conflicting ends; such an agent, in acting, both does and undoes his own projects in the same act. Conflicting desires point their bearer in different directions; one telling him to do one thing, the other telling him not to. Consistency pressures apply to such an agent; arguably, they forbid him from willing courses of action that are mutually exclusive. But it isn’t necessary to regard a basic desire as rationally impotent to be “of one mind” about what to will or how to behave. Unity can be had simply by ranking one of the desires lower in the order of preferences.
I don’t intend to challenge the subjectivist’s claim that a basic desire does not admit of intrinsic rational criticism. I accept it. In order for a desire to be intrinsically irrational, its irrationality must depend upon nothing extrinsic to the desire. In particular, the intrinsic irrationality of a desire must not depend upon any contingent fact about the person who has the desire or about the effects of treating the desire as rationally potent. According to my proposal, basic desires are ruled out by something extrinsic to the desire, not—notably—by the instrumental effects of having the desire, but by a contingent fact about the person who has it, namely, the contingent fact that the agent cares about a personal ideal that rules it out.

The Rejection of Categoricity

Subjectivists characteristically doubt the existence of categorical practical reasons and categorical norms of practical reason. This is a common way for subjectivists to distinguish their own tradition from its competitors, realism and Kantianism.

Subjectivists, of course, think that this doubt is well-founded, a theoretically virtuous instance of skepticism. The following chain of thought characterizes how many subjectivists think of this self-described strength. Robert Nozick describes subjectivism as the theory to beat. (Nature of Rationality, 133). Here are some of its strengths. First, it is generally accepted that practical reasons, as construed by subjectivism, are a part of the correct, fully-fledged theory of practical rationality; some basic desires do generate practical reasons: your thirst, for example (Sobel, “Pain for Objectivists”). Anti-subjectivists generally balk at the idea that only basic desires generate practical reasons, or the idea that all basic desires do—which, of course, is the topic of my project. Second, the means-end irrationality I described a moment ago is, most
everyone agrees, a clear example of practical irrationality. Any adequate theory will include it among its catalogue of types of practical irrationality, even if it is not accounted the only type.

Third, it is relatively clear why subjectivists think that practical reasons, as construed by their own theory, have a special type of normative force. The paradigmatic case of practical irrationality, in subjectivist accounts of practical reasoning, is an agent who fails to take the acknowledged means to an acknowledged desired end, absent any countervailing practical reason not to. The agent who fails in this way cannot plausibly argue that the practical reasons he failed to act upon don’t matter to him. As Hubin remarks,

Humeanism, unlike its competitors, connects reasons for acting with psychological states of the agent that are typically, and conceptually, connected to the agent’s motivational structure. When such reasons are presented to the agent, he cannot . . . respond that he doesn’t care about the evaluative perspective that grounds these reasons (“What’s So Special?”).

A person who doesn’t give a lick about being musical can certainly ask why he should care about the claim ‘You should practice your scales’. He doesn’t want the end to which practice is a means. The practical advice given to him is not connected to his own practical point of view, to something he wants. But within the subjectivist tradition, practical reasons are defined in terms of what the agent desires. They are, as described before, “agent-relative.” Hubin colorfully describes them as “unshruggable.”

Kantians and Aristotelian realists hold, on the other hand, that there are categorical practical reasons, or categorical rules of practical reasoning. Let’s focus, for the sake of clarity, on categorical rules. A categorical rule, in the relevant sense of ‘categorical’, is a rule that is

---

93 Sometimes this claim is construed in terms of a failure to be motivated to take an acknowledged means to an acknowledged basic desire instead of in terms of a failure to act upon one (Cullity and Gaut, “Introduction”).
valid for each rational agent, as such: an agent ought, as a matter of practical rationality, to follow it, no matter what he happens to desire.

We have to be careful about what we mean by ‘valid’. Normally, if I accuse someone of behaving immorally, then find out that he does not care about morality, I do not revoke the criticism. Moral rules “apply” to him, are “valid” for him, in a clear enough sense, even if he doesn’t care about them. A misbehaving person is, clearly enough, subject to moral criticism even if he does not care about moral ends. In fact, if he really doesn’t care about them, I am likely to add a stronger moral criticism yet. I’ll accuse him not of a momentary lapse in moral virtue, but a more significant character flaw. What the subjectivist means to be saying is that there aren’t any substantive rules that are ‘categorical’ in the sense of “rationally binding” upon an agent, no matter what he cares about.

For instance, Kant thinks that the Moral Law is a valid rule of rational willing for every rational agent, as such. A rational agent is rationally obligated to obey the dictates of the Moral Law, even if he doesn’t care about moral ends, and doesn’t care about any ends to which the moral life is a means. A categorical reason is a consideration that each agent ought to treat as valid for his practical deliberation no matter what he happens to desire. If, for instance, a human being is, in terms of human nature, a social animal, an Aristotelian realist might think that any human being, as such, has a practical reason to participate in society: the hermit who does not happen to have any desire-based interest in living with other people has, even so, a practical reason to do so, a categorical reason.

Subjectivists generally accept Bernard Williams’ well-known argument that realist and Kantian conceptions of non-instrumental practical reasoning which correct basic desires are often heavy-handed, dogmatic, and unconvincing (“Internal and External Reasons”). Also, it is harder
to see how categorical practical reasons, or categorical constraints upon practical reasons, given
that they are “agent-neutral,” are supposed to motivate an agent independently of what he cares
about. They seem “shruggable,” by conceivable agents, in the very least.

Norms of Rational Impotence, Personal Ideals, and Subjectivism

It might be helpful to locate my proposal within this discussion. I’ve said that I will
provide an account of the rational impotence thesis that is consistent with subjectivism. How so?

My proposal is consistent with the basic elements, as I argued earlier. It has the
“inclusive tendency” that subjectivist often display. I do not appeal to any form of intrinsic
rational criticism of basic desires, or to any categorical reasons or categorical rules, in the sense
of “rationally binding no matter what you happen to desire.” My theory regards a basic desire as
rationally potent so long as it is not rendered rationally impotent by norms of rational impotence
grounded in the agent’s ruling passions. Norms of rational impotence operate as “defeaters.” (A
basic desire is, as it were, “rationally potent unless rendered impotent.”) This preserves the idea,
central to subjectivism, that there is a standard and generally direct connection between having a
basic desire and having a practical reason to take the means to it.

Correspondingly, I don’t think that rational agency, as such, includes norms of rational
impotence. (This is one point of disagreement between Korsgaard and me.) For an agent that
does not have any ruling passions that include norms of rational impotence, each basic desire is
rationally potent. We could imagine a hedonistic egoist who cares about nothing, literally
nothing, except the maximization of his own physical pleasure, who doesn’t have any personal
ideals that include norms of rational impotence. We could stipulate that he does not find any pleasure in, or suffer any pains from, reflections upon what kind of person he is. In my view, he is an agent. He has desired ends and is able to employ his capacity for reason as a tool to bring them about. But he is not subject to any norms of rational impotence.

I also attempt to maintain the normative force of standard practical reasons for norms of rational impotence. Valid norms of rational impotence, as I construe them, are agent-relative, linked in a very tight way to the motivational states that centrally make up the agent’s practical point of view. According to my proposal, norms of rational impotence are aspects of a personal ideal that the agent cares about. They stand in a mereological relationship—a part/whole relationship—with the objects of the agent’s own desire.

At the same time, I don’t want to make the relationship too tight. I don’t deny that a person might no longer care about a personal ideal once he sees what it really asks of him. But this is true of the rational potency of garden-variety basic desires, as well. A person might want to become a pianist, realize how much practice it would take, and lose the desire. I submit—and this is all I submit—if an agent cares about being a good father, recognizes that this requires that he collect or weigh practical reasons in some way, and is not moved to reject the ideal, these requirements are valid for him. He can’t “shrug” them off; if he finds himself devoting time and energy to seeking out means to the satisfaction of what he recognizes to be a rationally impotent

---

94 Korsgaard would regard ‘hedonistic ethical egoism’ as a practical identity, and she points out that it does supply substantive practical advice about how to treat “the desire of the moment.” It does not follow, though, that it includes norms of rational impotence.

95 The example of the unwilling addict could be constructed so that the unwilling addict has a desire to treat his addictive impulse as rationally impotent. I suspect, in such a case, the desire to treat another desire as rationally impotent is not a basic desire that is “doing the normative work” of rendering the first-order impulse rationally impotent. It reflects, instead, another aspect of the agent’s psychology: acceptance of a norm of rational impotence, say.
basic desire, or he finds himself unmotivated to categorically reject it, he is guilty of a type of practical rationality that is akin, conceptually, to means-end irrationality.

A Loose End

There are, though, some interesting wrinkles to consider. First, I think that subjectivism shouldn’t overstate its case against “categoricity” in practical reasoning. In “Humean Doubts about Categorical Norms,” Dreier argues—successfully, I think—that subjectivists ought to admit that there is one categorical norm of practical reason, the instrumental principle. This principle is a valid rule of rational willing that a person ought to follow “no matter what he happens to want”—even if he does not possess a desire to follow it. An agent who has a desired end, recognizes a necessary means, but fails—without having a stronger, countervailing practical reason—to take it, is guilty of practical irrationality, of violating the instrumental principle. Were he to plead “I don’t desire to be means-end rational” that wouldn’t get him off the hook. If he cares about the end and recognizes a means to it, the rational thing for him to do is take it, regardless of whether he has a positive desire to be means-end rational.

Dreier argues that admitting that the instrumental principle is an objectively valid, universal, and categorical rule of rational willing is not grounds to be suspicious of the subjectivist theory. The instrumental principle is analogous to the rules of logic. Theoretical reasoning is in terms of valid forms of reasoning. Modus ponens transmits truth from ‘If A, then B’ and ‘A’ to ‘Therefore, B’, and this rule is normative for the reasoning of a person even if the person does not have a positive desire to reason soundly. Likewise, the instrumental principle transmits normativity from an agent’s rationally potent basic desires to adequate means. Practical reasoning is in terms of valid forms of thinking through practical questions. In this way, it
operates similar to the rules of logic. What subjectivists are committed to thinking is that there aren’t any other categorical rules of rational willing.96

What does Frankfurt mean by ‘Categorical’?97

In the introductory chapter, I pointed out that Frankfurt speaks of the will’s practical authority to “categorically reject” his violent impulse. This might create suspicion that he is implicitly rejecting his general commitment to subjectivism. He’s not, though. Frankfurt does not have in mind the sense of ‘categorical’ that Kant intends when he calls the supreme principle of morality, the “Categorical Imperative.” Kant means to be saying that the Categorical Imperative is a universally valid rule of rational willing. Given Frankfurt’s subjectivism, it is

96 Korsgaard first remarked that the subjectivist theory, in accepting an instrumental principle, includes a categorical norm of practical reason. She attempted to charge subjectivism with being a “companion in guilt” with the realist and Kantian theories it criticizes. I take Dreier’s argument to be an adequate response and accept the spirit of the following pithy remark: “Certain aspects of the Humean position deserve to be abandoned. We should abandon a hard-line metaphysical position according to which the very idea of practical reason is mysterious. Our skepticism should consist in doubts that the content of practical reason is anything like the content of morality, doubts, indeed, that its content outruns the bare [means-end] principle. We should be contesting the normative ground, not contesting its very existence” (45-46).

97 There is a question about why the concepts that Frankfurt attempts to analyze by way of the example of the unwilling addict and his present desire for heroin have proven so elusive. The answer is, in part, because the example is both over-described and under-described (in ways that, perhaps, many philosophical examples are). Many of the features of the example, as Frankfurt describes it, call out for theoretical understanding: his phenomenology, his disapproval of his own motivational states, his sense of personal defeat when he acts upon the addictive impulse, his suggestion that “he” is not participating in the process of taking the drugs, and his claim that he is “helplessly violated.” In effect, I have isolated one of the unwilling addict’s claims, his claim that his addictive impulse should play no role whatsoever in action, stipulated that this is what is central to “outlawing” the desire, and am focusing on how to understand it. The example is also under-described in the sense that Frankfurt tells us that the addict dissociates himself from his addiction. It strikes me that there are many senses in which an agent is able to dissociate from a mental state. He could be claiming that the desire is not consistent with his character, doesn’t fit into the common pattern that makes up his psychological history and profile, or isn’t relevant to judgments concerning praise and blame. Another explanation for the elusiveness is that Frankfurt’s theorizing, it seems to me, has undergone a shift of emphasis from his earliest work, which tends to focus upon questions about moral responsibility and the conditions under which an action is free, to his more recent work, which tends to focus upon questions about practical normativity and how to live. The rational impotence thesis is a thesis about practical normativity.
clear that his rejection of his violent desire isn’t on the basis of a norm that he regards as rationally binding upon every rational agent, as such. Its source is a ruling passion, which issues, in Kant’s terminology, “hypothetical” imperatives—commands that apply to him only in virtue of a desire he happens to have. However important this desire is to him, it is a contingent fact about his psychology.

It might be useful to think about what Frankfurt does mean. Let me rule out one common connotation of the term. Imagine that someone standing accused says, “I don’t merely deny the charges against me; I deny them categorically.” I hear two connotations. First, ‘categorical’ operates as an intensifier. It has, if used with the right tone of voice, an emphatic connotation. Second, it indicates that the accused won’t give an inch, no half admissions or concessions of guilt to weaker charges. I imagine that Frankfurt might well say the word ‘categorical’, in response to his violent impulse, with great emphasis, but I doubt this is a feature of the example that belongs as a part of the analysis of rational impotence. It shouldn’t be, in any case. An unwilling addict, weary of his addiction, might say “I have rejected this desire’s claim categorically” in an utterly flat tone.

Think about what it means to say that a moral rule is categorical. This use of categorical is close in meaning to ‘absolute’. A categorical rule against lying is a rule that is exceptionless. Frankfurt might well think that any violent impulse he has against one of his children is to be ruled out. Violent impulses, as a class, are rationally impotent. But this meaning does not seem to be primary. Frankfurt seems simply to contrast ‘categorical’ with ‘instrumental’; his violent impulse is not merely to be ruled out on instrumental grounds, it is to be ruled out wholesale and without dependence upon instrumental considerations.
Think of it this way. Imagine that you have come to think that the moral rule against lying is absolute. Acts of lying aren’t to be treated as *prima facie* wrongs such that any instance of lying should be treated as wrong unless an adequate justification is present to indicate that this particular lie is an exception to the general rule. There aren’t, you think, any exceptions. This flies in the face of common moral attitudes. Lying is often considered a *prima facie* wrong. Some people think that lies that save another person from embarrassment or hurt feelings are permissible: “white lies.” Others think that more significant moral ends must be served for a lie to be permissible, an exception to the general moral rule. However advocates of a general but not absolute rule against lying see this issue, they agree that it makes sense to ask, of any instance of lying, whether it is morally wrong, all things considered, or whether it is merely a *prima facie* wrong. They then would naturally search around to see if any of the exception-granting conditions happen to apply.

If lying is morally wrong in each case, it follows that standard moral considerations—saving people’s feelings, saving their lives—will not speak strongly enough against the rule to create exceptions to it. This has practical implications: it would take away the need to consider, in practical deliberation, whether there are countervailing moral considerations that speak against acting following the rule. There is no need to undergo these reflections, at least not to attempt to weigh them against the merits of obeying the rule; the rule is categorical, and the act the rule forbids is not merely generally wrong, it is always wrong.

Perhaps what Frankfurt wants to say fits this pattern. Treating a violent impulse against one of his children as reason-giving is absolutely wrong—it is never permissible. This violent impulse doesn’t have any normative weight whatsoever. Consequently any reflection upon the various elements of his circumstances, which might otherwise be relevant to his judgments about
what to do, can be ignored—even if his wife wouldn’t find out, he wouldn’t be sent to jail, would be able to take a pill that helped him forget, so that he didn’t suffer from the pain of the shame and guilt that would otherwise overwhelm him. His impulse is ruled out independently of these considerations. It doesn’t take these consideration to beat it or even to weaken it. The desire is ruled out, wholesale, independent of these standard considerations. It is ruled out, we might say, “on non-instrumental grounds” and “without qualification”—without qualifications, in any case, such as “if my wife might be nearby” or “if my neighbor can overhear.”

In a subjectivist theory, practical reasons are grounded in contingent, subjective motivational states. Consequently, there is a “condition” attached to practical reasons, and so to judgments of rational impotence. Norms of rational impotence are hypothetical in Kant’s sense. But in the event that the ‘if’ clause is satisfied, the agent does have the relevant desired end, the imperative is valid. And if it is shaped like a norm that rules out some basic desires, does not admit of exceptions, and is insensitive in the way I’ve described to considerations that are often valid, it is—in a sense—a categorical demand.

It is in virtue of one of Frankfurt’s desires that the impulse is to be categorically rejected. A norm of rational impotence is valid for me so long as I have a ruling passion that grounds it: so long as I care about what I do. If my practical point of view were to change so that I no longer had a ruling passion that grounds a norm of rational impotence, it would no longer be valid for me. If Frankfurt were to come to have a different desiderative profile, were he to become a person who does not care about being a decent father, then the norm of rational impotence would no longer be valid for him.98 In some sense, it could come to be the case that a basic desire to kill

98 Here is a curiosity. For a time, I intended to construe the categorical rejection of a basic desire as a form of non-instrumental, non-intrinsic rational criticism. Should I? It does operate differently than a standard
his child is rationally potent for him, but it would have to be a sense of ‘him’ that is consistent with a radical change in his practical point of view.

Last Remarks

Here’s what I’ve accomplished, as I see it, in this chapter. As well as laying out the terms of debate and indicating which protagonists I will argue with, I’ve argued for the claim that the rational impotence thesis, properly understood, is consistent with both the basic tenets of subjectivism and with several of its characteristic ideas, which I take to be central to the theory. By itself, this isn’t a positive reason to accept the rational impotence thesis. It is a reason to think that the rational impotence thesis is a candidate for inclusion. I have been playing my hand, on
occasion, giving some sense of how it fits with the ideas that drive subjectivism. But in the next chapter, I want to do more. I want to give my fellow subjectivist her chance to speak; then I’ll indicate why a subjectivism that advocates the rational impotence thesis is better.
CHAPTER 3
WHY IT'S BETTER TO BE A DISCRIMINATING SUBJECTIVIST

Reason . . . ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them. (Hume, *Treatise*, 415)

Should a subjectivist about practical reasons accept the rational impotence thesis?

I think so. I think there are norms for rational impotence, with subjectivist groundings.

Imagine, though, that a fellow subjectivist disagrees. Unlike me, she thinks that every (logically satisfiable) basic desire is rationally potent. What we’d have is a family dispute.

This domestic quarrel has theoretical consequences. It means, among other things, that I and my fellow subjectivist respond differently to the anti-subjectivist challenge. On the one hand, I accept the rational impotence thesis and reject the incompatibility thesis; on the other, she rejects the rational impotence thesis, and so—whether or not she thinks this thesis is *logically inconsistent* with subjectivism—her theory, in either case, *doesn’t* countenance the existence of rationally impotent basic desires. Which branch of subjectivism is preferable?

My primary aim in this chapter is to indicate why subjectivists should take my side in this internecine spat. In chapter 2, I argued that a thorough-going subjectivist is able to affirm, with

---

99 Some subjectivists have expressed this idea: for instance, Hubin (“Irrational Desires”), though he changes his mind (“Desires, Whims, and Values”).

100 From this point, for the sake of a fluidity, I will drop the qualification ‘logically satisfiable’. Any subjectivist theory is able to account for the rational impotence of logically unsatisfiable basic desires. Since nothing could count as a means to the satisfaction of a logically unsatisfiable desire, it can’t generate any practical reasons.

---
logical consistency, the rational impotence thesis. But this, by itself, isn’t a positive reason for subjectivists to incorporate the thesis into their theory. There need to be reasons to. At the end of this chapter, I will give some.

Taking my side might seem counterintuitive, initially. My fellow subjectivist’s theory begins with a theoretical advantage. It’s simpler. I bother to distinguish between rationally potent and rationally impotent basic desires and thereby incur a theoretical debt. If I have any desire to escape the charge of incompleteness, I have to account for the distinction by making sense of the rational filtering of basic desires: how are some basic desires rendered rationally impotent?101 My opponent regards every basic desire as rationally potent; she makes no appeal, explicit or implicit, to rational filters or substantive rules for the collection of practical reasons.102

Also, the tension in Frankfurt’s account also exists in mine: it isn’t immediately clear how the will, construed in subjectivist terms, has the capacity to disenfranchise basic desires. As I put it earlier, how do desires rule out desires?

The debt is payable and the tension dissolvable, I think. This is an instance in which greater theoretical complexity is a virtue: the pursuit of what you most care about might well compel you—rationally—to disenfranchise some basic desires that well up in the run of life.

The family troubles needn’t get ugly. Frankfurt claims that it is “preposterous” to deny the rational impotence thesis, and he is very critical of (unnamed) subjectivists for operating with a “meager conceptual repertoire.” But I don’t think that subjectivists who deny the rational impotence thesis are incapable of responding in interesting and sophisticated ways to the anti-subjectivist challenge. I will consider, rather extensively, how they could. As a subjectivist, I

101 Hubin speaks of a “value screen” for the evaluation of basic desires (“Desires, Whims, and Values”).

102 Other than, of course, ‘If you have a basic desire and an adequate means, you have a practical reason’.
have a vested interest in bringing to light all of the theoretical resources of the Humean tradition
to counter the anti-subjectivist challenge. As a matter of fact, I will adopt into my own fully-
fledged response to the anti-subjectivist challenge much—though not all—of what I say on behalf
of subjectivists who disagree with me.

With respect to my fellow Humean theorists, then, I come as a peace-maker—mostly. I
went to some length in the previous chapter to reveal that the truth of the rational impotence
thesis does not threaten the central tenets of subjectivism. My proposal doesn’t appeal—
explicitly or implicitly—to realist standards of objective value or Kantian rules of practical
reason, as such, or anything similarly robust. A legitimate appeal to a norm of rational impotence
isn’t heavy-handed or dogmatic in the ways that an appeal to norms of practical reason proposed
by anti-subjectivists often are. My subjectivist account posits norms that are agent-relative,
grounded in the agent’s own ruling passions, the pursuits and ideals he himself cares about.
Accordingly, my account secures for legitimate norms of rational impotence the
“unshruggability” that standard subjectivist practical reasons enjoy; the agent who violates the
norms of rational impotence that are rationally binding upon him is failing to align his behavior
with his own practical standards. In means-end irrationality, the voice being ignored by an agent
is the voice of one of his own desires; the same goes for an agent who violates a valid norm of
rational impotence.

I do, though, dig in my heels and insist on one point—family peace be damned. My
account puts into better focus what I take to be one of the most central insights of the subjectivist

---

103 An agent could conceivably care about personal ideals that encourages him to be tough on himself, of
course. But if he points out to himself, or another points out to him, what his personal ideal demands of
him, the authority of the demand is grounded in elements of his own subjective motivational set. It is not
indepent of, and so is not insensitive to, them.
tradition. Let’s get at it through reflection upon Hume’s famous quotation, in which he says that reason ought only be a slave to “the passions.” Which passions?

Hume’s remark might lead someone to think that subjectivism is motivated by the belief that an unmotivated desire is, by its very nature, intrinsically normative, “born to rule.” A basic desire, as such, is a “king” with the “divine right” to command reason to seek out means to its satisfaction. Or, given how many basic, unmotivated desires the average person has, perhaps we should choose a more democratic political metaphor: a basic desire, as such, is to be regarded as a “citizen” that ought to be treated with dignity and respect, which amounts in the very least to giving it a voice in practical deliberation. Such a subjectivism advocates, we might say, a “great democracy of desire.” (Plato would accept the metaphor, frown at the theory.) In this way of thinking, all passions—or, in any case, all basic motivational states—get to “rule,” and simply in virtue of being basic desires.

I don’t see things that way. First, I think there are basic desires that don’t deserve a voice in practical deliberation, so I’m not a thorough-going, “universalistic” democrat. I reject some basic desires. (Not to mention, I’m no egalitarian, either, I privilege some desires over others.) Second, I don’t think that desires are to be regarded—akin to persons—as intrinsically deserving of respect. The thought that provokes me to accept subjectivism isn’t that desires are special and, on analogy with human beings, “born” with various inalienable rights such as “the right to be listened to.”

---

104 As is common with slogans, Hume’s is multiply ambiguous and vague. What does being a “slave” to the passions mean? Does reason merely seek means? Or does it have other duties? I think that reason does more than seek means; for instance, it also clarifies ends and helps the agent discover ways to make contingently conflicting ends fit into, in the very least, a moderately coherent life plan.
Some things are dear to our hearts, and to act rationally is to look after these things, as best we can. An agent might care deeply about the welfare of her children, or a social ideal, or some other state of affairs. In my view, “reason” ought to serve, first and foremost, these cares and concerns. It is a “slave” to ruling passions, not to passions generally. For a person who does have ruling passions, things dear to her heart, many of her weightiest practical reasons derive, ultimately, from them, and any valid norms of rational impotence (as well as other practical policies and plans) are grounded in these, too. These norms, policies, and plans gain the right to dictate how an agent ought to act and even to deliberate by being properly related to his ruling passions. Reason ought to seek means to them, clarify what these passions actually demand, consider how to fit various cares and concerns into a coherent life plan, and so on.

Naming Names, Giving Titles

It would be useful to give names to the competing species of subjectivism. One option is to continue with the political metaphor and call them “democratic” and “undemocratic” subjectivism. A citizen within a democracy deserves a voice within the political process, among other things, the right to cast a vote in elections. Likewise, a subjectivist who rejects the rational impotence thesis consequently accepts that each satisfiable basic desire generates a practical reason and so deserves a voice in practical deliberation. We might say, naturally enough, that she believes that each basic desire has a “vote” within the process of deciding what to do.

This political title could be misleading, though; and it’s useful to point out why. In a democracy, a citizen has the right not only to cast a vote, but to have that vote given the same

---

105 A anti-subjectivist might go the other way and claim that a desire, as such, never generates practical reasons: for instance, Scanlon (What We Owe Each Other, 50-55).
weight as any other vote cast by any other citizen: each vote counts for one, and no vote for more than one. That is, with respect to the voting rights of a citizen, democracy is generally both universalistic and egalitarian. But rejecting the rational impotence thesis only commits a subjectivist to universalism, and she is likely to find egalitarianism implausible. It is a very common and very plausible intuition that some basic desires consistently generate, for a particular agent, especially weighty practical reasons. My desire for the creamy texture of ice cream, I think, is rationally potent, but it doesn’t deserve a seat at the table as important as my desire to spend time with my children. Subjectivists tend to accept the idea that agents have weightier practical reasons to pursue what they desire more, even if they don’t appeal to valuings and ruling passions to make sense of this intuition. (I labor this point because I intend to exploit it.)

Another proposal is to call those who accept the rational impotence thesis either “restrictive” or “discriminating” subjectivists and those who don’t “non-restrictive” or “indiscriminate.” Those who accept this rational impotence thesis place restrictions upon which basic desires are to be included in practical deliberation. Discriminating subjectivists believe that there are basic desires that do not have what it takes to deserve a voice within deliberation. Non-restrictive, indiscriminate subjectivists, on the other hand, make no restrictions and do not assess credentials (other than ‘being a basic desire’). Though hardly vivid, these titles have the virtue of being clear.

If I were to prefer flavorless names, the ‘restrictive’-‘non-restrictive’ labels would be preferable. To call someone ‘indiscriminate’ is commonly an insult, as it suggests that he is boorish, insensitive, and lacking standards. But the non-restrictive subjectivist, to be clear, rejects the existence of the relevant standards, and he might be right. If so, his “lack of discrimination” is a virtue, not a vice: it is not a weakness to reject a standard that does not exist. Whether
legitimate standards for the collection of practical reasons exist is one of the questions we’ll be examining in this chapter.

My decision: for the sake of clarity, I will generally use ‘indiscriminate’ and ‘discriminate’.

The Agenda

The primary question this chapter addresses is whether discriminate or indiscriminate subjectivism is preferable. Which theory better fits the evidence and the legitimate desiderata for a theory of practical reasons?

Though I intend to structure this chapter in terms of a squabble between fellow theorists in the Humean tradition, I want to keep in mind what the neighbors will think. I don’t want to lose sight of the relevance of this internal debate to the long-standing dispute between subjectivists and their opponents. Subjectivists want, of course, to show well when they confront anti-subjectivists, and I will suggest that the best hat to wear—the theory that is best-suited to challenge competitors—is a subjectivism that includes the rational impotence thesis and several corresponding concepts: personal ideals and norms of rational impotence. My proposal not only coheres with subjectivism, it accounts for intuitions that might otherwise seem to be positive evidence for anti-subjectivism.

First, I am going to explore the indiscriminate subjectivism’s response to the anti-subjectivist challenge. Later, I am going to say more about why I think the greater theoretical

---

106 There are, of course, questions about which species of subjectivism is rhetorically stronger. But I’m not concerned with rhetoric. I mean to be seeking out which theory better tracks the evidence and better meets legitimate desiderata for a theory of practical reasons.
simplicity of indiscriminate subjectivism should be eschewed in favor of restrictive, discriminating subjectivism.

The Arguments for the Rational Impotence Thesis

So how might an indiscriminate subjectivist argue against the rational impotence thesis? I mentioned two arguments for this thesis in chapter 2: the argument from the experience of practical deliberation, which includes an appeal to purported examples, and the argument from full reflective agency. She might begin by attempting to undermine them.

The argument from the experience of practical deliberation first states that it is coherent to ask, of any particular desire, whether it generates a (pro tanto) practical reason, and it would be striking if the answer were always ‘no’. Moreover, don’t we already recognize, pre-theoretically, that some basic desires are rationally impotent? We classify some desires as reasons; we discount others altogether. From strange, fleeting impulses to persistent temptations, from quirky whims to horrific wants, we experience basic desires that we are averse to treating as reason-giving. Why not conclude, the argument might say, that what seems to be the case really is? Why distrust our intuitions about what counts, and what doesn’t, as a legitimate source of a practical reason and our common way of classifying desires? The argument concludes that we shouldn’t distrust these intuitions: the purported examples are legitimate examples; there are, in the very least, some rationally impotent basic desires.

The suggestive examples I described in the previous chapter give us plenty to talk through. The question, then, is whether any of these purported examples is a legitimate example of rational impotence.

---

107 Cohon, “The Roots of Reasons.”
What is the Conclusion?

One aspect of the anti-subjectivist challenge that is unclear is the conclusion. The challenge could aim at conclusions of varying strength. For instance, at any of the following: subjectivism (a) is (demonstrably) false; (b) is (highly) improbable; or (c) suffers from an objection that is serious, though not, by itself, theory-killing. The version of the argument that aims for the strongest conclusion—(a)—states that the conclusion—‘subjectivism is (demonstrably) false’—can be established validly from two true premises; the rational impotence thesis and the incompatibility thesis are both (demonstrably) true, and they jointly imply that subjectivism is (demonstrably) false. But, other versions of the argument, with weaker conclusions, are also conceivable. If our best arguments suggest that the rational impotence thesis and the incompatibility thesis are both very likely true, it would follow that—(b)—there is strong inductive evidence against the truth of subjectivism. Or, if the rational impotence thesis appears, after full reflection, to be true and it is a mystery how subjectivism would account for it, then (c) follows—there is a hole in subjectivism: there is a plausible thesis that the theory hasn’t accommodated, an appearance that it hasn’t saved. This argument, if sound, would favor—to one degree or another—the rejection of subjectivism. The seriousness of the challenge would depend on how large the hole is, and that depends upon how central or important to practical rationality and practical deliberation the standards for rational impotence are. For that mystery to approach a “theory-killer,” the standards would need to be central or important. In any case, it will be important to keep in mind which conclusion is being argued for, and how the subjectivist could best respond to the argument understood in that way.
How should the indiscriminate subjectivist respond? Let’s separate the argument into two parts. One part of the argument is the appeal to purported examples; the other is an appeal to intuition, or perhaps to “pre-theoretic” intuitions: people sometimes “discount a desire altogether,” and there is simply no good reason to deny that these discountings are sometimes legitimate. I would like to handle each of the two parts of the argument separately, beginning with the appeal to intuition.

The Appeal to Intuition

The appeal to intuition is not sufficient, by itself, to get to the strong conclusion that the rational impotence thesis is (demonstrably) true. There are reasons to distrust the relevant intuitions.

First, Cohon’s appeal to intuition—“Don’t we already discounts some desires altogether”—grants a high degree of trust to a particular type of experience. It suggests that we ought to judge that what seems to be the case really is. But our gut feelings, it seems, are not likely to track the correct answers to complex questions about the purpose and function of practical-reason-giving and the nature of practical reason. I don’t think we have, at our immediate disposal, all the information relevant to this debate.

Cohon seems to treat her intuitions as “pre-theoretic.” Are they? Or do they reflect aspects of her considered theory? It’s not uncommon for a so-called “pre-theoretic” intuition to

---

108 In many contexts, we ought to balk at this inference. We might hear similar pieces of reasoning, with a suspect history, “It seems that when I am acting, I am acting freely. Therefore I must actually be free.” Wittgenstein tells a story intended to call this inference into question. As the wind blows a leaf to and fro, it says “I think I’ll go this way now; and now I think I will go that way.” The point of the story is that the leaf feels free, but this does not entail that it is. The leaf doesn’t recognize that the wind is the causal force determining its direction. Failing to recognize the causal role of the wind is not the same thing as being causally free from it. The point: we might not have introspective access to the relevant causal chains.
be an expression of a reflective attitude rather than a thought that guides us to them. If the appeal to intuition were nothing more than an expression of the author’s anti-Humean convictions, the appeal is question-begging.

Also, it’s not immediately clear that “discounting” a desire is equivalent to judging it rationally impotent. Cohon speaks of how “we” classify some desires as reasons, and not others. People do step back from their own desires and experience strongly negative reactions to them. And in some instances, people denounce or dissociate from desires they do not like. I agree, that is, that there is a class of events that requires analysis, especially if we describe things at a very generally level. For instance, what parent hasn’t suffered a flare of temper and recoiled in horror at the sense of aggression she feels within herself? We might hear her say, “That gives me no reason to attack my child.” The language of “having a reason” is in every day life. But it is harder to imagine an every day agent who says “This basic desire of mine, though there are effective means to its satisfaction, does not generate any pro tanto normative consideration for me.” The language of ‘basic desire’, ‘rational impotence’, and ‘pro tanto practical reason’ are technical terms; they are not every day words.

The every day denouncings that the argument is gesturing at require interpretation, and admit of several. It is conceivable, for instance, that the act of discounting does not reflect the judgment that the desire in question is rationally impotent, but that it is clearly so weak it will not win out in competition against other practical reasons. It is not that the desire deserves no recognition whatsoever; it is simply that a moment’s reflection makes clear how weak the reason it generates is, and so a moment’s reflection is all it deserves. This denunciation is “post-collection” weighing, not “pre-collection” rejection.
Also, the subjectivist might attempt to provide various debunking explanations. The belief that a desire is rationally impotent could be an instance of wishful thinking.\footnote{To illustrate, a spectre hangs over belief in human freedom. The suspicion is that many people who believe in it do not have good evidence for the belief; they simply want so badly to believe it. Wanting something to be true, however badly, is no evidence that it is. Similar suspicions might arise about judgments of rational impotence.} The loving mother despises her violent impulse, and it is understandable that she does not want to cede it any authority, any voice, whatsoever. But not wanting a citizen with unseemly political affiliations to have the right to vote does not mean he doesn’t have it.\footnote{I don’t, by the way, think that wishful thinking is, in every instance, without virtue. The parent who convinces herself that her missing child is doing alright might be guilty of wishful thinking, but in so believing display an admirable resolve or a sensible form of self-protection. Likewise, the mother who is unwilling to cede any authority to her violent impulse, even if she happens to be mistaken, might well be fetchingly feisty, but mistaken.} Likewise, not wanting a desire to have a voice is not itself grounds for thinking that it doesn’t have it. There have to be proper grounds for its extrusion.\footnote{Good thing, too. If I happen to be in a mood I generally despise, and I find myself wanting not to care about my children any more, the theory that not wanting to treat a desire as rationally potent means it isn’t would render my desire to help my children impotent, for the amount of time I’m in the mood.}

A strong denunciation of a desire could also be performative. Judging her violent impulse rationally impotent could be a way for the loving mother to assure her husband, or herself, that she does not intend to act upon it, or to assure him, or herself, that she really does care deeply about her child. It isn’t unusual for a person to look for evidence in herself, among her own psychological states, that she really does care about what she thinks she does. A violent impulse could be seen as a bit of counterevidence.\footnote{I don’t think that it necessarily is. There are mental analogues to twitches, which say little about us.} Hoping to maintain her sense of herself, she denounces it.
These alternative interpretations,\textsuperscript{113} whether true or false, do fit the phenomenon. If Cohon means to suggest that most or all people do have the intuition that there are rationally impotent desires, that strikes me as overstating the phenomenon. That takes an argument, not a mere appeal to intuition. And even if such judgments are widespread, they could be systematically mistaken.

I don’t want to put too much stock in debunking explanations. Debunking explanations tend to have the most force when it is difficult to explain how a thesis with so little evidence is so widely accepted or so persistently held, and even when the paucity of evidence is recognized. The interesting question, though, when discussing the truth of a thesis, is whether there is evidence, and whether it is sufficient, and we haven’t worked our way through all of that yet. So, for the time being, take these debunking explanations as nothing more than a vivid demand to be provided with that evidence, as well as attempts to short circuit the argument from the experience of practical deliberation. Until we have better evidence for the rational impotence thesis, it is appropriate to wonder whether whatever attraction the rational impotence thesis holds is more about our hopes and desires than about the evidence.

The appeal to intuition also turns on the general idea that it is coherent to ask, of any particular basic desire, whether it is rationally impotent. As discussed in chapter 2, the indiscriminate subjectivist does not need to deny that this. The idea of a standard by which to judge desires rationally impotent is, in one sense, clear enough. We understand in the very least what the function of the standard would be: it is a collection principle that forbids treating certain basic desires as proper input into practical deliberation. But many coherent standards bear no

\textsuperscript{113} For a discussion of different types of negative reaction to a desire, see Arpaly and Schroeder (“Alienation and Externality”).
practical authority whatsoever. A traffic rule that requires driving on the left side of the road is coherent, but not valid in the United States.

The anti-subjectivist argument expresses that it would be striking were the answer to the question ‘no’ in every single case. The indiscriminate theorist might agree that such a result is striking; he merely is committed to saying it is true. Not to mention, there are contexts in which an appeal to a counterexample is sufficient, by itself, to reveal that a thesis is false. I say to you, “I’m not so sure there are any prime numbers smaller than 13.” You say: “What about the number 3.” I then see that 3 is smaller than 13 and that it meets the definition of ‘prime’: “not being divisible without remainder by any pair of whole numbers other than 1 and itself.” Once you produce the number 3 and we’re clear about definitions, I have to admit that you are correct; it would be pure obstinacy not to. But in the case concerning prime numbers, we know what we are looking for, what the criteria for ‘being smaller than 13’ and ‘being prime’ are, and we can check the purported examples to see whether they meet the criteria. In the debate about the rational impotence thesis, on the other hand, an appeal to purported examples is not enough. Though it is clear what it means for a desire to be rationally impotent, it is not clear what the criteria for rational impotence are.

At this stage of our discussion, we lack criteria, or anything approaching criteria. It might be too much to demand a full-fledged account of rational impotence before we are to judge that the purported examples are (likely to be) legitimate examples. We don't yet have an account of such a thing, so it’s not clear whether the purported examples are legitimate examples. Certainly, denying that they are legitimate doesn't involve a theorist in the same level of obstinacy as denying that 3 is a prime number less than 13. Again, absent at least a rudimentary
account, it is hard to see why we should think that these (pre-theoretic) intuitions are to carry much weight.

So, taken as an attempt to argue for the strong conclusion, the appeal to (pre-theoretic) intuition is unsound. In all fairness, this aspect of the argument might be nothing more than an attempt to shift the burden of proof. But there are serious objections to the appeal to intuition even on this weaker construal. The author of the argument has his own theoretical burdens. Absent an account of the standard by which desires are purportedly to be judged rationally impotent, the defender of the rational impotence thesis is a companion in guilt. We are aware of proposals that include practical standards that generate practical reason claims that do not bear any practical authority. The anti-subjectivist has several important questions to answer, too. Without an account of the “impotent-making features” of the purported examples, his argument bears, at best, a kind of *prima facie*, provisional plausibility. The subjectivist does, after all, give an explanation for his central tenet that basic desires do not admit of intrinsic rational criticism: that’s the point of the talk about how desires are goal-directed and not truth-directed states. And the subjectivist will throw his own suspicious glances at the realist and Kantian: shouldn’t we be suspicious that these norms will be unconvincingly heavy-handed or dogmatic?

There’s no reason to continue in the present direction. I’ll draw a conclusion: I don’t see any special reason to think that the burden of proof is on the indiscriminate subjectivist. And I do not intend to attempt to weigh out who has the greater burden. It strikes me that both the advocates and the opponents have them: the anti-subjectivist defender of the rational impotence thesis needs to give an account of the relevant standard(s) that desires are to be judged from; the

---

114 Kant, of course, does have a significantly-developed account. I will discuss it in chapters 4-6.
opponent needs to give an account of why the purported examples are not real examples.

Everybody has work to do.

This is how to respond to the first thrust in the argument. But what about the appeal to purported examples? That is, after all, likely to be the stronger part of the argument.

The Appeal to Purported Examples

The indiscriminate subjectivist, by definition, rejects each of the purported examples of rational impotence. Instead of simply declaring that each example proposed by her opponent is illegitimate and leaving it at that, she could further her cause by explaining how she would treat the various purported examples. If she is able to interpret the purported examples of rational impotence in a way consistent with her own subjectivist theory, and then argue that her analysis of the various purported examples is equally, if not more, plausible than the interpretation of rational impotence, then she will have gone a long way towards blunting the anti-subjectivist challenge.

I am going to consider an array of strategies, under three headings. There are an indefinite number of conceivable examples, and I won’t discuss all of them. The idea is that different purported examples are going to be interpreted more plausibly by different general strategies. And I’ll leave it to the reader to apply these strategies with wisdom and discretion. Also, to be clear, my discussion in this section is of hypothetical examples, I have no intention to attempt to “settle” any real life cases. The task at this point is conceptual: how could subjectivists analyze the purported examples? I’ll generally leave it to the psychologists to tackle the empirical question concerning what’s really going on in people’s heads.
A judgment of rational impotence is a complex judgment. If a person judges that a particular mental state is a rationally impotent basic desire, she commits herself to the following claims: (1) the mental state in question is a desire, (2) it is not only a desire, but a basic desire, and (3) the appropriate standard by which to condemn it is a standard of rational impotence. I will suggest that each of these judgments could be mistaken. And in the process I will describe the underlying complexity of the standard “trumping model” to which subjectivists commonly appeal.

(1) Distinguishing between Different Types of Criticism of Desire

The first general strategy has to do with the third claim. The indiscriminate subjectivist might argue that, at least in some cases, a purported example of rational impotence is a basic desire that ought to come in for some type of reflective criticism, but that the appropriate type of criticism is not intrinsic, rational criticism.

Instrumental, Rational Criticism

One version of this general strategy makes a distinction between intrinsic rational criticism and instrumental rational criticism. Consider the unwilling addict with a present desire for heroin. Let’s stipulate: he cares deeply about his health and loves his children, both of which would suffer were he to act upon his addiction. The addiction bears a degree of causal potency that is far greater than the degree of practical authority it has; when it wells up, it tends to subvert his better judgment and his stable intention not to take the drugs—and he knows it. If he treats his desire for drugs as rationally potent, even weakly so, he has a far more difficult time fighting it than if he takes a more strongly negative attitude toward it. Give it an inch, and it takes a mile.
The addict then has an instrumental reason not to give his addiction a seat at the table. To take an attitude of utter disrespect, as it were, to a lesser, trumped desire might play an important causal role in keeping it in its proper place. In this case, it does. This constitutes an instrumental reason—grounded in his basic desires for health and the welfare of his children, and the relevant facts about how to secure them—to treat his basic desire as though it were rationally impotent. But having an instrumental reason to treat a basic desire as though it were rationally impotent is not the same thing as its being rationally impotent.

The instrumentally-motivated denunciation of a craving, exhibited by the addict, mimics more direct, deontic criticism in that it rules out treating the desire as rationally potent, and before the process of weighing begins. But the source is not a standard for rational impotence, it is an instrumental consideration, grounded in the agent’s own subjective concerns. Moreover, are we to think that most agents are sufficiently sensitive to this distinction to warrant the judgment, which Cohon seems to make, that their strong condemnation of a desire is an intrinsic, and not an instrumental, rational judgment?

To augment this general strategy, the subjectivist might make a distinction between ‘having’ a practical reason and ‘being under a requirement’ to give that practical reason a voice in practical deliberation. To this point I have spoken as though ‘being rational potent’ and ‘deserving a voice’ are the same thing. This strategy suggests that the truth is more complex.

To illustrate, consider the following example. I have a standing basic desire to experience pleasure, but there are things that I care about more than pleasure. The basic elements

\[\text{\footnotesize 115 Some of our subjective concerns are of such importance that we ought to commit time and energy to keep a watchful eye against desires that tend to subvert them.}\]
of the subjectivist theory seems to imply that I have a practical reason to do any of the very large number of things that would give me even the smallest tidbit of pleasure. Given that it would seem that a consideration is given a voice in practical deliberation only if I give it, in the very least, a rudimentary amount of thought, it might seem as though I am under a requirement to give a thought to doing each of the things that would bring me pleasure. To put this metaphorically, we might imagine the process of weighing reasons in as a process of writing down of each reason and its comparative weight and so on. To count as having treated a consideration as a reason, I would have to actually get the consideration down on the sheet and find reasons against it that are stronger. To discharge from the metaphor, I would have to do, in my mind, whatever happens to count as writing down and weighing in the metaphor. I haven’t, though, taken stock—made a mental note—of all of the practical reasons generated by my basic desire for pleasure. And I haven’t compared all the possible means against each other. To say the least, we rarely, if ever, go through the process with this level of care and precision and thoroughness. (How tedious.) In my case, there are abundant means, and I blithely ignore many of them.

As I see it, I ‘have’ these practical reasons, but I’m not guilty of practical irrationality for failing take all of them into account. I have, we might say, practical reasons of rational advice to take these various means, but not in each case a corresponding practical reason of rational virtuosity to pay attention to them. An ideal rational advisor—who is omniscient—would be aware of them; it’s not a failure in my performance as a practical reasoner that I am not.

The model of practical deliberation I have been working with is idealized. The ideal of decision-making that our discussion suggests is that, in making a decision, an agent ought to collect together the various practical reasons he has—all of them—and then weigh these various practical reasons against each other to come to a final ‘ought’ of practical reason. The ideal tends
to suggest that every practical reason be collected and weighed in. But, no doubt, agents commonly do not meet this ideal, and we hardly should expect them to, or to accuse them of practical irrationality if they don’t. There is a difference between a standard for the correctness and a practical policy for real life decision-making.

There is also an instrumental justification for this commonplace inattention, a generic justification which applies to most any agent no matter what he happens to care about, whatever his most cherished pursuits. Even if the indiscriminate theorist is committed to thinking that there is a standard connection between having a practical reason and that consideration deserving a voice in practical deliberation, there seem to be cases in which a countervailing instrumental reason severs this standard connection. Human beings have limited time and energy, and weighing more considerations has costs in terms of them. It’s not that instrumental considerations render the basic desire rationally impotent; they simply absolve our lack of attention.\footnote{Consider the debate about maximizing and satisficing. A maximizer is a person who aims to take the best possible option, of those available. A satisficer takes the first option that is satisfactory. Satisficing, in effect, cuts down the number of options evaluated, and so saves time and energy. Clearly, for human beings, who have only so much of these “commodities,” satisficing is often the preferable option. This goes to the point that it is not, in itself, irrational to ignore some practical options.}

Also, sometimes there are other costs—to our sense of self, for instance. A loving mother who gives a voice to her violent impulse against her child might feel a painful sense of betrayal for doing so, in which case she has an instrumental consideration not to.

Criticism from a Substantive Practical Standard

Second, it is also important to be clear about the type of reason claim subjectivists are talking about. Subjectivism is a theory of the considerations relevant to rational advice and the rational appraisal of action. A ‘practical reason’ is a consideration that an agent ought, on pains
of rational criticism, to give weight in his practical decision-making. These reasons are valid for your practical decisions, and—according to subjectivism—in virtue of your own desired goals.

We speak colloquially, though, of various kinds of reasons: moral, prudential, political, economic, and so on. Hubin distinguishes, helpfully, between practical reasons and ‘reasons from a point of view’ (“Prudential Reasons”). There are reasons of etiquette, which derive from a system of rules that together form the rules of etiquette. Etiquette is a substantive point of view that generates, in one sense, reason claims. You have a reason, a reason of etiquette, to say ‘thank you’ when you’ve been given a gift, for example.

There is a clear sense in which reasons of etiquette are applicable to any person. We are able to apply the rules of etiquette to any person’s circumstances and see what practical advice, if any, it generates. There is nothing mysterious about these rules, and nothing mysterious about the fact that a substantive set of rules generates reason claims of this sort. But there is also a clear sense in which reasons of etiquette do not apply to everyone, and could conceivably fail to have any normative force for an agent. An agent might notice what etiquette tells him to do and have no practical reason to do it. According to subjectivism, whether the rules of etiquette, and so reason claims of etiquette, are rationally binding upon you depends on your desired goals, your actual motivational states. If you don’t care about etiquette, and following the dictates of etiquette isn’t prescribed by anything you do care about, then the reasons from this point of view are not practical reasons for you, and you would not be irrational were you not to give them a voice.

This second sense of reasons are practical reasons. Practical reasons are in answer to ‘why should I do that?’ questions, too. But as Cullity and Gaut put it, “The practical reasons that
answer the root ‘Why should I do that?’ questions” contain “a should for which” further requests for practical justification are unnecessary (“Introduction,” ix). Or, as Dreier puts it,

There is a fundamental sense . . . of having a reason . . . To think that there could be reasons to follow the rules of rationality is, I would say, to misunderstand what reasons are. Reasons are in terms of the rules of rationality (“Humean Doubts,” 29).

According to subjectivism, ‘Why do that?’ questions end, properly, in appeals to basic desires.

Subjectivists do not think that prudence or morality supply practical reasons for any rational agent, as such. (This is controversial, of course; but not among subjectivists.) Prudential and moral reasons are practical reasons for an agent, valid considerations that he ought to take into account in practical deliberation, only if he cares about them, or cares about something that requires him to treat them as valid. It is unusual but conceivable that an agent does not desire his own welfare. That prudential reasons are valid for so many human beings is that most people do care about their own welfare. Likewise, moral reasons are practical reasons for an agent only if morality is properly linked to ends the agent does care about, an intrinsic concern about moral ends, or something else.

Let’s apply these ideas. A second version of this general strategy distinguishes between intrinsic rational criticism and intrinsic criticism from a substantive normative standard. Take Cohon’s example of the hateful desire to stick it to an unfriendly colleague. She claims that this impulse fails to generate a practical reason; it seems she regards it as rationally impotent because the desire is immoral. I agree that her hateful desire—an instance of icy, self-protective, professional envy, I take it—does not give her any moral reason to stick it to the person she hates. From the moral perspective, the desire is a temptation, not a legitimate consideration to be included along with other moral considerations in making the final judgment about what morality prescribes. Within moral deliberation—thinking about what morality asks, or even demands, of
us—we ought to factor into the calculation only those considerations that are morally legitimate, and a hateful desire is not. But moral deliberation is not, in itself, rational deliberation.

Similar objections might be made to the example of the pacifist who judges there is no reason whatsoever to strike back. The pacifistic norm is not, the indiscriminate subjectivist might point out, a norm of rational impotence, but a norm that belongs to the pacifistic ideal, which is not rationally mandated.

The distinction between practical reasons and perspective-based reasons for acting is common within the Humean subjectivist tradition, which does not regard substantive practical standards such as the moral perspective as intrinsically rationally binding. It isn’t hard to imagine cases in which the reasons that derive from a substantive practical standard fail to have any binding rational authority over an agent. A practical standard is rationally binding upon an agent only if the agent has the appropriate subjective concerns. The indiscriminate subjectivist could ask for an argument why we ought to think that the moral standard, or any other substantive practical standards, are standards of rationality. (In following chapters, I am going to examine Kant’s and Korsgaard’s attempts to argue that the Moral Law is rationally binding upon every agent, as such.) What is needed here, then, is more than an appeal to an example of a morally questionable basic desire.

117 No doubt, some versions of desire-satisfaction utilitarianism count the mere fact that someone wants something as a positive moral reason. But this is, in itself, one of the most serious objections to desire-satisfaction utilitarianism. It should factor out immoral desires.

118 For a clear discussion, see Hubin (“What’s So Special?” and “Desires, Whims, and Values”).
(2) Mistaken Interpretations of Mental States.

As mentioned, a judgment of rational impotence is a complex judgment. If someone judges that a basic desire is rationally impotent, he is claiming at least three things: the mental state coming in for judgment is a *desire*, that it is a *basic* desire, and that it fails to meet a standard for rational potency. In the previous section, I consider ways of arguing against the third claim. A second general strategy, of which there are several versions, is to suggest that the agent who experiences what she regards as a rationally impotent basic desire is misinterpreting—not the judgment, but—her mental state. People do make false mental state attributions, even to themselves; we misinterpret the blooming, buzzing confusion of our own mental lives. Perhaps some of the purported examples of rational impotence reflect mistaken mental state attributions. The mental state in question is not a desire, or is not, in any case, a basic desire.

Under-described Desires

Consider a teenage athlete misinterpreted by his doting mother: “I want a trophy on my shelf that says ‘High School State Champion’ in golden letters,” he says. His mother makes it so: she buys him a shiny, embossed trophy and places it next to the myriad of pictures she has adoringly placed on his shelf. The athlete’s real basic desire is not satisfied; what he really wants is to *earn* a state championship, to win one on the playing field. Buying one does not cut it, and so the mother has no practical reason to pay out the cash.\textsuperscript{119}

Apply this idea to the present debate. In some cases, such under-description could be the real culprit behind a judgment of rational impotence. Consider the pacifist who is shoved unprovoked in the run of a pick up basketball game claims to feel a purely vengeful impulse. But

\textsuperscript{119} See Hubin (“Irrational Desires” and “Desires, Whims, and Values”).
a desire to respond to a gratuitous push need not be “pure” or “mere” vengeance. Perhaps, what he really feels is warranted anger or righteous indignation at the (relatively minor, but no less real) injustice of his fellow player. It is not that he wants to hit back; he wants to put things aright. It is not that he wants to strike back for the sake of striking back; he wants—appropriately, as I see it—to re-assert his own value: no one deserves to be treated in the way that the overly aggressive player has treated him. He cares about victims; he wants to see any victim’s value re-asserted after it has been ignored. In this case, he happens to be the victim.

If this interpretation were accurate, whether a vengeful impulse would be rationally potent for a pacifist is beside the point. If there is no vengeful desire, there is no vengeful desire to disenfranchise. To continue, his real basic desire gives him a practical reason to do what would most effectively re-assert that value. And that is not, from his point of view, an objectionable desire. In this case, what appropriately comes in for criticism is not the basic desire from which the judgment ‘I have reason to strike back’ derives; it is the means. Shoving back is not the most effective means of re-asserting value; it is too easily mistaken for mere vengeance and is not likely to get across the moral point. A more appropriate expression of moral judgment would be, perhaps, to eject the offending player from the game, or to threaten to eject him unless he atones for his mistreatment of the pacifist by an apology, or something like that. In any case, striking back is off limits, and it is unlikely to be the most effective way of expressing the appropriate moral point, anyhow.

I think this strategy does have application to real life cases. It is not uncommon for people to describe their own mental states less than thoroughly, and it is not uncommon for people to describe them mistakenly. In the case I’ve discussed, the pacifist has described his
desire less sympathetically than is called for—whether in a spirit of humility, on the basis of an overly pessimistic theory of human nature, or on some other grounds.

A Mistaken Judgment of Basicity

A second version of this strategy claims that a purported example of rationally impotent basic desire is not really a basic desire. A desire is basic just in case it is not based upon another desire, along with means-end beliefs.

If we were confronted with a real life example of a mother who professes to have experienced a violent urge against her beloved child, we might suspect that the impulse she experiences is not, in reality, a basic desire to harm or to kill the child. It would not be surprising for a loving mother, if her child has been crying for some time, to want—in a basic way—peace and quiet: for the child to stop crying. In an uninhibited piece of means-end reasoning, the mother recognizes that drowning the child is a (very morbid but) expedient means to her basic desired end. Desired end hooks up with acknowledged means, and the mother finds herself with a motivated (instrumental) desire to drown the child. Before she has time for more self-conscious reflection, she finds herself leaning in toward the child.

No doubt, without any attention to the finer points of a discussion about its etiology, the mother is likely to be shocked by her impulse, and what it is asking her to do. And she is likely to distance herself from it almost immediately. But it would not be surprising to find the mother wondering where the impulse “came from.” She seems to conclude that she is feeling hostile toward her child. But that interpretation is under-motivated. Along with the characteristic connection between desire and action and desire and affect, we might learn of a desire by its triggering a bit of means-end reasoning.
To say the least, on plausible assumptions about the mother’s various desired ends and her particular circumstances, killing her child is gruesome and grisly in its effectiveness, but inconsistent with a number of other concerns that she has: she doesn't want to lose her child; she doesn't want to feel the guilt of having killed her child or to face her husband; she doesn't want to go to jail; and so on. On this very natural and sympathetic construal of the mother’s actual psychological constitution, she is mistaken about the character of her own desire. She is taking the desire to kill her child as basic, though it is not. And, in this case, the desire for peace is basic and the desire to drown the child is motivated. No doubt, there are other ways of getting the peace, many of which incur fewer and less grisly costs. She could call a sitter.

The question naturally arises whether there are very many such examples in real life. But this strategy does seem to have wide application. Recall the happy person who feels a strange desire to veer into oncoming traffic. The whim is shocking and calls out for understanding. But if the person were to wonder whether it is a basic desire, we would have reason to be skeptical. The suicidal whim could be triggered by curiosity: “What would it be like to do something so dangerous?” Or—to appeal to an idea attractive to some existentialists—by a fit of existential freedom, a rebellion even against our “most sensible” selves or our most cherished values.

A desire is characterized, in many cases, by a push or a pull. And there are questions, in each case, about the character and source of the push or pull. Is this a basic desire? Is it motivated? From what basic desire, and what means-end belief, does it derive? And so on. The indiscriminate subjectivist could argue, for many purported examples, that we don’t have adequate reason to think the strange impulse is really a basic desire.

Here is a related issue. Perhaps there is a tendency, in some people, to interpret their own desires as being not only basic but “deep” in a way that, I think, many of our desires are not. I
don’t have a clear description of this idea, but some forms of romanticism encourage us to learn who we are by getting in touch with our “innermost being,” as though the “real self” sends dispatches to the “conscious self” from deep within. (Does Freud deserve some blame?) The loving mother might feel a violent urge and infer that she bears a deep (and shocking) hostility toward her child. Her recognition of the desire leads her to feel shame, but this shame is unwarranted. She is not hostile; she wants the child to stop crying, and that desire triggered—in a way insensitive to several of her other basic desires—a piece of means end reasoning that, if followed, would have awful consequences. She’s frustrated, not hostile. Temporarily overwhelmed, not permanently and deeply antagonistic.

One of Freud’s patients, Rat Man, might have made this kind of mistake. Rat Man suffered from ambivalent feelings toward his father. He both felt affection for, and hostility towards, his father. Rat Man’s personality subsequently split. He would speak, at times, as though the person who hated his father was a distinct person, not really him: a distinct locus of responsibility, with a distinct set of character traits and beliefs and so on. On one interpretation, Rat Man interpreted each of his competing desires as the expression of a deep self. Since he had conflicting feelings, he interpreted this as the source of conflicting deep selves. He then built a conception of identity around each desire, and then flitted back and forth between the “person” who loved and the “person” who hated his father. Did Rat Man have good reason to read into natural feelings, either of affection or hostility, a “deep” reflection of his innermost being? I

---

120 It isn’t unusual for a person who from time to time feels, say, envy to speak of “my envious self” or “the envious person in me.” I find this practice, generally, very innocuous. It’s a shorthand way of speaking about patterns of feeling that we discover in the course of our psychological history.

121 Velleman raises this example as an objection to Frankfurt’s theory (“Identification and Identity”). He regards Rat Man as an extreme example of a general problem for any theory that posits a “real” self.” I think that Velleman misinterprets Frankfurt, though this is a tricky issue and I won’t get into it here.
don’t know enough about Rat Man’s case, in particular; but I’m suspicious of the practice.
Feelings of hostility do admit of alternative explanations, and if Rat Man favored his feelings of affection, he could have rendered his hostility normatively weak from the perspective of this preference (or vice versa).

So, perhaps the mother’s interpretation of her violent urge as a basic desire, or even as a reflection of her innermost being, it might be nothing like that. It is a motivated desire, based upon an unobjectionable basic desire, and a true, but highly objectionable, means-end belief that, if acted upon, would have severe practical costs. It is the means that deserves practical criticism.

Off-Limits Means

Perhaps the example of the loving mother does suggest an interesting constraint on the indiscriminate subjectivist theory of practical rationality. The subjectivist theory accepts an instrumental principle. According to the instrumental principle, as I’ve described it, an agent has a (pro tanto) practical reason to do whatever counts as an effective means to whatever rationally potent basic desires he has. The instrumental principle communicates normativity from the rationally potent basic desire to the means to its satisfaction. We could treat this as a first approximation and ask whether there are exceptions to this standard rule. Perhaps some effective means are to be rejected.

Again, the loving mother is likely to think that she doesn’t have any practical reason to harm her child. It is consistent with this judgment that either a basic desire to harm the child is rationally impotent for her or the action ‘harm your child’ is something she could not have a practical reason to do no matter what that action might be a means to. Whether parents ever consider it or not, children cost money, and most people prefer to have more rather than less
money. Does a basic desire for a better financial future generate, for anyone who prefers more money to less, a practical reason to dispense with the child? I would think that most parents would dismiss the suggestion, not because they do not really want a better financial future, but because they are unwilling to entertain the actions that count as adequate means. These actions—killing, passing off, or abandoning their children—are “unthinkable,” utterly off limits.

So, another possible interpretation of the mother’s rejection of the practical reason claim ‘you have a pro tanto practical reason to kill your child’ is that it is a rejection of the means, not the basic desire. She is unwilling to treat any action that can be appropriately described as 'killing my child' as a practical option. This would imply that the mother is committed to a basic volitional norm that functions, from case to case, deontically: it rules out a particular type of behavior as a practical option.

I think such volitional norms are relatively common. For instance, according to a traditional ideal of marriage, adultery is not a practical option for the married. No matter what benefits would come from acting in this way, this type of behavior is ruled out by the behavioral norms of the agent’s marital ideal. It’s a part of the vows. If the action-type cannot be rightfully enacted, given the personal commitment, it rules out treating the action-type as a practical option. The behavioral prohibition extends to the process of practical reasoning: “Don’t worry, honey, I have a practical reason to betray you, but I’ve promised, remember, not to act on it” doesn’t seem to cut it. Arguably, this type of prohibition is a feature of the pacifist’s ideal, as well. It is a central feature of this ideal that vengeful actions are not to be done. Such behavior is not an option for him.

If we were to accept this qualification, the instrumental principle should read: “You have a practical reason to do whatever counts as an effective means to the ends that you have a
practical reason to bring about, other things being equal’, where the *ceteris paribus* clause reads “unless you are rationally committed to a norm not to treat the activity that counts as an effective means as a practical option.” In which case, the marital ideal and the pacifistic ideal and, in the example of the mother, the parental ideal includes norms for volition, in particular, norms that forbid treating particular kinds of behavior as legitimate practical options. Call an action that, in a mere descriptive sense, counts an effective means to an end desired by an agent, but is ruled out by a volitional norm to which an agent is committed an “off limits” means. Perhaps the mother’s violent urge is a motivated desire, generated by her rationally potent basic desire for her child to stop crying and an off limits means.  

Mistaken Desire-Attributions

People can be mistaken not only in thinking that a desire is basic but in thinking that a mental state is a desire.

Think of the happy person’s “urge” to veer into oncoming traffic, or his vertiginous “desire” to step off of a steep cliff. Don Hubin suggests an interesting way to think about these examples. In some instances, perhaps the person who experiences a “suicidal whim” is not really experiencing a desire. He is, instead, merely “entertaining” one. With respect to beliefs, people often “try on for size” a proposition that they do not accept. Philosophy professors are familiar with asking students to grant a claim for the sake of argument and reason from it. This “thought experiment” proceeds as if the person does hold the belief. Without believing the proposition,

---

122 See Frankfurt ("Rationality and the Unthinkable").
inferences are made from it, evidence is considered for and against it, and to the practical effects of so believing are thought up and weighed.

Likewise, a person might try a desire on for size. It doesn’t take a desire to provoke means-end reasoning. I might imagine what it would be like to want to kill myself, then go through the process of thinking about what it would feel like to want this or what it would be like to act on such a desire, while never having the desire: there is no push or pull characteristic of a desire. In this case, my mental state is a mere ‘imagining that’; it is not a motivational state. If this is what is really happening, then the indiscriminate theorist has any easy way to handle the case. The judgment that my suicidal “whim” is rationally impotent would be mistaken: there is no desire present, and so no desire to disenfranchise.

(3) The Complexity of the Trumping Model

One way to see the disagreement between subjectivists is as a disagreement about the appropriate model for practical deliberation. Any subjectivist is likely to think that, in some cases in which there is a conflict between competing basic desires, each of the various competing desires generates a pro tanto practical reason. And, in many of these cases, one of these practical reason is simply “stronger” or “weightier” than other, competing practical reasons, that is, that one practical reason “trumps” the others. The indiscriminate theorist thinks that every case of conflicting desires can be assimilated in one way or another to the trumping model. Her opponent, on the other hand, thinks that there are cases that the trumping model cannot handle. (When Frankfurt gestures at the existence of rationally impotent basic desires, he has often made a distinction between two types of conflict between desires.) But how could the indiscriminate Humean argue that the purported examples can be adequately assimilated to the trumping model?
One idea that might be behind the general challenge to the Humean theory is the thought that the trumping model is simply crude and simplistic. Think for a moment about the trumping model. The basic idea behind the trumping model is that some practical reasons are outweighed by others. For instance, if I find in myself both a desire to go to my sister’s wedding and a desire to spend the day relaxing at home, but I can’t do both, then I have a practical reason to leave home and go to the wedding and a practical reason to stay at home and miss the wedding. It is natural to say that, on plausible assumptions about me, the practical reason generated by my desire to go to my sister’s wedding is “stronger” or “weightier” or “trumps” the practical reason generated by my desire to stay at home. Whatever practical reason I have to miss the wedding, I have a stronger practical reason not to.

The basic idea here is simple, I suppose. But the indiscriminate Humean might respond by attempting to reveal complexities behind the superficial simplicity of the trumping model. If he is able to do so, he might then exploit these complexities to do the work that the silencing model is supposed to do.

This strategy requires that we reflect upon the language commonly employed by the trumping model. Along with the idea of one practical reason trumping another, it is also common for theorists to speak of one practical reason being “stronger” than another, or one practical reason being “weightier” than another. These terms are metaphorical. The talk of ‘weight’ suggests that there are scales to weigh the practical reasons against each other. The idea of trumping comes from card playing, a King is worth more points than a nine and so “trumps” the nine. And the concept of strength here is not the idea of brute (causal) power, but normative right. An executive might not have the greatest power to make good decisions, and yet have the
authority to make them. These terms—weight, trumping, and strength—are metaphorical, but they are natural metaphors.

Let’s think about the concept of weight for a moment. I take it that there are at least three ways of thinking about the weightiness of a practical reason. One thing, though, which we need to keep clear is whether we are speaking of the weightiness of a practical reason or of the basic desire that generates it. It seems that important basic desires, in some cases, generate relatively “weak” or “light” practical reasons. Your basic desire for good health, I take it, is important; but in some cases this basic desire, along with an effective means, generates a pro tanto practical reason without much weight. You have a practical reason, in terms of your desire for good health, not to eat that donut. But, since the health effects will be relatively minor, the practical reason is relatively weak. Perhaps, the fact that eating the donut would be enjoyable is reason enough to warrant doing so, whereas your basic desire for good health gives you a very strong reason not to drink poison. So, an important basic desire does not always generate a weighty practical reason, and it will be important to keep this in mind. It’s easy to slide between talking of the importance of a basic desire and the weightiness of a practical reason it generates.

Also, how should we speak of the following possibility? If two basic desires each generate a practical reason to do one and the same thing, should we say that you have two practical reasons to do one and the same thing, or one practical reason with two sources? I will

---

123 It is also common for the terms ‘weight’, ‘trumps’, and ‘strength’ to be defined, if they are defined at all, in terms of each other. Whether or not it can be done, I am not going to attempt to get out of this circle of definitions. Instead, I want to think about whether these simple metaphors conceal any complexity that the indiscriminate theorist can exploit.
speak of you having two practical reasons to do one and the same thing. The reason why I choose to speak this way will be clear in a moment.\textsuperscript{124}

Head-to-Head Comparison

One sense of weightiness has to do with “head-to-head” comparisons between competing practical reasons. If an agent has (say, two and only two) competing basic desires in a particular choice situation, we might wonder which desire generates, by itself, a weightier or stronger practical reason. That is to say, if we were to isolate the each of the practical reasons that the two basic desires generate, line these two practical reasons up side-by-side, and weigh them against each other, which would weigh more? If we were to place each on opposing sides of a scale, which, if either, would push the scale down?

Clearly, there are cases in which one practical reason, generated by a single desire, is stronger than another practical reason, generated by another single competing desire. For instance, the practical reason to go to the wedding generated by my desire not to offend my family wins out in a head-to-head competition with the practical reason not to go generated by my fleeting and trivial desire to take a long nap. I care deeply about family relations; I merely have a sleepy whim. In head-to-head comparison, the practical reason to go trumps the practical reason not to. Here I don’t mean to deny that it is possible for two practical reasons to be (roughly) equal in weight; I merely mean to assert that in some cases, one is stronger in this sense than

\textsuperscript{124} There is also the possibility that one and the same basic desire could generate competing practical reasons. A desire to exercise might generate a practical reason to jog and a practical reason to swim, when there isn’t time for both. I won’t discuss this possibility. I’ll consider cases in which distinct basic desires are in competition.
another: it beats it head-to-head. One sense in which a practical reason is (relatively) weighty is that it wins in head-to-head competition with (most) other practical reasons.

Ganging Up

A second sense of weightiness has to do with several basic desires “ganging up” against a single, competing basic desire. Going to my sister’s wedding is supported by several basic desires. First, I have an intrinsic desire to experience the wedding, for which going is a necessary means. Second, my family would be hurt if I were to stay home, and I do not want to offend. Third, I would deeply regret missing the big event, and I have no desire to suffer. And so on. The desire not to go is fleeting and trivial; and so, the practical reason it generates is not only beaten, individually, by each of the three considerations in favor of going, it is resoundingly beaten by the group of practical reasons generated by my various other basic desires. Here, a number of basic desires corporately generate a particularly strong set of practical reasons to go to the wedding.

The concept of “ganging up” is especially important, of course, in those cases in which a practical reason that is strong in head-to-head comparisons is outweighed by a group of individually weaker considerations. In these cases, the possibility of ganging up makes it possible for an individually strong consideration to lose out in the battle to be the final practical reason. The ganging up of weaker considerations changes the decision. Though the possibility of ganging up does not change the outcome of the process of weighing in every case, it is useful to the Indiscriminate Humean, for it also makes possible instances in which a singularly weak practical reason is beaten resoundingly by a corporation of singularly strong practical reasons. And that is what happens in my case.
A third sense of weightiness has to do with “jurisdiction.” A general has greater authority than a corporal in the sense that the general’s orders win in head-to-head comparison. In a case in which the general and the corporal issue conflicting orders to a mutual underling, the general’s orders trump. But the general also has command over a greater number and range of people and has a greater range of prerogatives than a corporal. In short, a general has greater jurisdiction. He is able to give orders not only to privates and to majors, but also to corporals. He is also able to order operations that the corporal is not permitted to order.

To carry this idea over to the trumping model, one basic desire might be weightier than another in the sense that it has the authority, as it were, to generate a greater number of practical reasons than the other. My desire to take a long nap gives me, the indiscriminate subjectivist might say, a practical reason to satisfy it, but it gives me no more than that. It does not also deserve, for instance, the right to command my faculty of instrumental thinking to seek out appropriate means to get out of the wedding: a manufactured cold, say. My various basic desires that each generate a practical reason for me to go to the wedding do, however, also give me a practical reason to seek out a map to the church, a practical reason to begin to consider what I will wear, a practical reason to brush my teeth, and so on. My desire not to offend, then, has greater jurisdiction, in this case, than my desire to relax. Some basic desires generate a practical reason, but nothing more than a practical reason to satisfy the basic desire. Other basic desires generate a wider range of practical reasons—to engage in further practical reasoning, for instance.

As Bratman might put it, some basic desires ground intentions that function as partial plans that will require further thinking and filling in later. Such a desire has the authority, so to speak, to command that the future self do whatever work is necessary to fill in the incomplete
parts of the plan when the time comes. But there are also some basic desires that simply fail to support any such intentions or plans. This makes sense of the talk that a desire sometimes “subverts” an agent’s will and “takes over” her mind. The unwilling drug addict’s desire, for instance, does not deserve to command that agent’s processes of instrumental reasoning, and yet it does cause her to seek out a means to the satisfaction of the addiction.

Application to the Purported Examples

Distinguishing between each type of weightiness might help the Indiscriminate Humean to generate a threefold strategy for interpreting (some of) the purported examples. In the first sense, weightiness is a matter of being able to win in head-to-head combat. And this type of weightiness seems to come in degrees. In some competitions, one practical reason is slightly weightier than another; in some cases, one is quite a bit weightier; and, in some cases, why not think that one practical reason is significantly or exceedingly weightier? It is far weightier in head-to-head comparison.  

Weightiness in this sense comes in degrees; and so, plausibly, it comes in a high degree. In the second sense, if a particular course of action is supported by a large number of considerations, whereas a competing course of action is supported by only one, rather weak consideration, the first group of practical reasons are resoundingly stronger in a corporate sense. In the third sense, a (group of) practical reason(s) might command greater authority to set off the process of instrumental reasoning, or to govern conduct generally.

---

125 I am not saying what about a particular desire grants it greater or lesser head-to-head weightiness, and I am not proposing how to assign such values. I have not supplied a precise way to measure respective weightiness. But, that aside, even if we remain at an intuitive level, it seems fair to say that some practical reasons are a bit weightier than others, some are quite a bit weightier, and perhaps some practical reasons are—it is more appropriate to say—(to borrow a term from Hubin) “swamped” by another practical reason. That is to say, weightiness in terms of head-to-head comparison comes in degrees, and so it is able to come in a strong degree.
If there are three ways for a basic desire and its practical reasons to be weighty, it seems there are three ways for them to lack weightiness. Why not think, then, that some of the purported examples of rational impotence are really basic desires that generate a pro tanto practical reason that is weak in each of the three senses? It loses resoundingly in head-to-head competition with each of a corporation of practical reasons that conflict with it, and so it does not command any authority to generate further practical reasons, whereas its competitors do. The Indiscriminate Humean might then conclude that judgments of rational impotence are, strictly speaking, false; but that does not mean that they do not have a kernel of truth: they are exaggerated expressions of the compound weakness of the practical reason in question.

Which of the purported examples fits this general, threefold strategy? Perhaps the happy person’s suicidal whim is best handled this way, though I have my doubts. Even if the whim gives the agent a practical reason to satisfy it, it gives him nothing more. It is individually weak, beaten corporately, and has no jurisdiction. His desire to live generates for him a practical reason to stay on his side of the road, and this desire is itself a central and important subjective concern. His competing suicidal whim simply does not generate a practical reason that is strong enough, in head-to-head competition, to defeat his practical reason to stay on his side of the road. Also, the driver not only desires to live, he desires so many things that require that he stay alive: he wants to benefit his children, he wants to finish his personal projects, he wants to watch the game on TV tonight, and so on. So, the whim generates a practical reason that is individually weak, and that is beaten by a corporation of other concerns he has, some of which are individually strong. Moreover, his whim does not generate any practical reasons other than the practical reason to
satisfy it. That practical reason, or the basic desire that generates it, is weak in each of the three senses so far mentioned. No doubt, it is fair to say that it is "swamped."126

Actually, I am not sure about the third part of this judgment. It seems to me that this type of whim might well generate a practical reason, not to kill oneself, but to engage in a piece of imagining. At least in some cases, a suicidal whim is, I think, an expression of curiosity: "What would it be like to be in a crash?" or "How would I react if I were about to die?" or "What would the world be like without me in it?" In these cases, the curiosity generates, I think, a reason to lose oneself in a daydream (though it might be best to wait until one gets out of the car), but not to actually steer into traffic. Insofar as the whim reflects this kind of existential curiosity, I think it does generate practical reasons. If not, if the whim is akin the strange instinct of the lemming, then I think it should be handled in the way I just suggested.

In summary, the first general strategy, which has three components, attempts to reveal that the trumping model is more complex than one might initially think. And this is supposed to permit the Indiscriminate Humean to interpret some of the purported examples of rational impotence as strictly speaking false, but with a kernel of truth: the basic desire in question does generate a practical reason, but an exceedingly and multiply weak practical reason. A judgment of rational impotence is false, but it is (in some cases at least) close to the truth. The

126 Or, it might be especially plausible to say that the strange whim to stick your finger in a gooey substance is best met by this strategy. This whim gives you a practical reason to stick, but there is no other practical reason to do so. And if it would make you nauseous, or force you to spend the day, uncomfortably and inconveniently, with goo residue on your finger, there are countervailing reasons strong enough to defeat the practical reason you have to stick your finger in. This presumes, of course, that sticking one’s finger in a gooey substance has nothing going for it. I think it has a kind of spontaneity, and it strikes me that the gross and the disgusting do have their odd allure.
Indiscriminate Humean might then encourage us to interpret the practice Cohon gestures at in this way rather than as evidence for the rational impotence thesis.  

127 This general threefold strategy to overcome the appeal to purported examples of rational impotence is structurally similar to a common defense of hedonistic utilitarianism. On one construal—Jeremy Bentham’s, for instance—the utilitarian theory is committed to hedonism. Hedonism is a theory of value. It is a complex thesis that states that (a) all pleasure is intrinsically valuable (or worth pursuing), and (b) only pleasures are intrinsically valuable (or worth pursuing). Hedonistic utilitarianism states that the morally right thing to do is the action that maximizes pleasure, everyone considered. What makes the right action right is that it is the action, of those available to the agent, that creates the biggest bundle of pleasure, every pleasure counted in. The theory provides criteria for moral rightness. And hedonism implies that any pleasure brought about by a particular action counts in favor, morally speaking, of doing that action. The first aspect of hedonism claims that every pleasure is intrinsically valuable. This implies that even what would generally be regarded as immoral pleasures count in the calculation that determines which action is morally right. Given its commitment to hedonism—in particular, to (a)—utilitarianism suffers from the criticism that it regards sadistic pleasures as having positive moral status. It commonly strikes people that sadistic pleasures can't possibly have such status, and any theory that implies that they do must be false. Just as the strict advocate of hedonism must admit that the pleasure experienced by the sadist counts in favor, morally speaking, of his sadistic behavior, this pleasure is swamped by the pain he causes his victims, and so the admission is not as damning as it might initially appear. It is not a “theory killer.” The defense of indiscriminate Humeanism I have proposed bears a structural similarity to the defense of hedonistic utilitarianism. Likewise, even if indiscriminate Humeanism implies that desires such as the mother’s violent impulse and the pacifist’s vengeful impulse are rationally potent, it does not imply that they bear a high degree of weightiness. They are outweighed, perhaps even strongly outweighed—that is, swamped—by the practical reasons generated by the agent’s other subjective concerns. A threat remains to the Utilitarian theory, though. The sadist might become a utility monster; it might turn out that the sadist gets so much pleasure from his pain-causing that utilitarianism implies that his sadistic behavior is morally right. That is, that morality would prescribe his sadistic behavior! If you grant the sadist’s sadistic pleasures into the moral calculus, those pleasures might acquire what it takes to carry the decision. Similarly, the threat exists for the indiscriminate Humean theory that the mother’s violent impulse or the pacifist’s vengeful impulse or some other unseemly and unwanted desire acquire a sufficient degree of whatever happens to be the relevant potency-generating feature. This is what fuels, I take it, part of the argument from full reflective agency, namely, the part that warns giving over authority to a fickle master. If you give it a bit of authority, circumstances might come about that would grant it final authority. I'm going to make something of this later. In summary, the first general strategy, which has three components, attempts to reveal that the trumping model is more complex than one might initially think. And this is supposed to permit the Indiscriminate Humean to interpret some of the purported examples of rational impotence as strictly speaking false, but with a kernel of truth: the basic desire in question does generate a practical reason, but an exceedingly and multiply weak practical reason. A judgment of rational impotence is false, but it is (in some cases at least) close to the truth. The indiscriminate Humean might then encourage us to interpret the practice Cohon gestures at in this way rather than as evidence for the rational impotence thesis.
A Final Analysis

This concludes my defense of indiscriminate subjectivism against the appeal to purported examples. More or less, what I have been arguing to this point is that there are a number of grounds upon which the indiscriminate subjectivist could reject a purported example, and a number of ways she might attempt to re-interpret it. She could place her hope in the idea that each purported example is, in the very least, handled at least as plausible by one or more of her deflecting strategies.

The Argument from Full Reflective Agency

Consider the argument from full reflective agency, which has a distinctly Kantian flavor. Our capacity for reflection is constitutive of, or at least makes possible what is constitutive of, our very humanity or personhood. What sets human beings apart from other animals is our ability to step back from our mental states and ask questions—in particular, normative questions—about them. And this makes possible, as Kant might point out, the activity of willing.

A fully rational human being asks a number of normative questions, both theoretical and practical. As for theoretical questions, he wonders whether his beliefs are true, and whether he has (sufficient) justification for them. He wonders whether he must have evidence for each of his beliefs, or whether there are some beliefs that are properly basic. He wonders whether there is anything he knows with certainty, and if so, what. Also, a fully rational being asks practical questions. He wonders how he ought to live and what he ought to do. He wonders what his various pro tanto practical reasons are and how to weigh them against each other; and he seeks out effective means to his desired ends. To stop such reflection short of asking whether the
various (basic) impulses we experience really give us practical reasons is to reject a question that stems from the very capacity that makes up our humanity.

The argument could also be put in terms of autonomy. An agent who is very good at instrumental reasoning, but whose values—whose ruling passions—are simply and unreflectively the consequence of his upbringing, is not fully autonomous. Full autonomy requires that we govern ourselves, and that requires, minimally, that we make our values “our own” by considering and endorsing those values for reasons. Failure to call into question the practical authority of our basic desires would, the argument might proceed, constitute heteronomy; it would be to give over authority to something foreign to our wills. Desires can lead us to action, but if the action is to be autonomous, the desires need to be endorsed by the agent.

How ought the indiscriminate subjectivist respond to this argument? First, if the suggestion is that the Humean theory is committed to discouraging agents from asking questions about rational impotence, that’s mistaken. It does not discourage the question; it answers it. If an agent asks, of a particular basic desire, ‘Is this desire rationally potent?’, the indiscriminate theory says ‘yes’. Second, it strikes me that asking this particular question is not a requirement of practical rationality, especially not of rational virtuosity. I do not think that an agent is guilty of practical irrationality if he does not ask meta-theoretical questions about the ultimate source of his practical reasons.

Strangely, Korsgaard denies this. She thinks there is a “rule of reflection” that requires us to seek answers to any questions reflection is able to raise. I would think that a better way to talk about the pursuit of theoretical understanding of the ultimate source of normativity is by distinguishing between an ideal of practical rationality, on the one hand, and the ordinary requirements of practical rationality or rational virtuosity, on the other. To behave rationally, to
be “rationally virtuous,” agents need only meet the ordinary requirements of practical rationality. Generally speaking, this includes some level of epistemic virtue—identifying ends and seeking means. It includes being means-end rational and—arguably—being sensitive to some consistency pressures. If she finds in herself conflicting pursuits, she will have to find ways to resolve them. But to be practically rational, she need not ask questions about those requirements, let alone have satisfactory answers to them. I prize a person with a Socratic temperament, who seeks out answers to the questions that arise from our every day practices, but I don’t charge non-philosophers, generally, with irrationality when they don’t “press reflection as far as it will go.”

Frankfurt speaks of a type of practical deference to desire that characterizes the lives of small children: “They do whatever their impulses move them most insistently to do, without any self-regarding interest” (“Taking Ourselves Seriously,” 6). At some point in the process of maturation, we “disrupt ourselves from an uncritical immersion in our current primary experience [and] take a look at what is going on.” Even if not in an “especially thoughtful” or “entirely overt” way, we begin to take attitudes to the “raw psychic material” that we discover in ourselves through reflection. A person might, if sufficiently thoughtful, disrupt her tendency to treat her various desires as reason-giving, and wonder, about one or other desire, whether she should. This is a move in the direction of greater maturity, but is it necessary for rational agency?

Third, the most serious objection to this argument is that it turns on a controversial conception of autonomy. Korsgaard claims that when you deliberate, it is as though there is something—you—that stands in judgment of your practical options and which chooses between

---

128 Scanlon talks about how at least many of our desires are “seemings”; they strike us as reason-giving, at least initially. An important part of reflective agency is to treat “seemings” as “problems, not reasons,” and “assess” whether they ought to be treated as reason-giving (*What We Owe Each Other*, 51; and “Reasons and Passions”).
them. This does track, I suppose, our every day way of speaking, but it doesn't follow from this bit of ordinary language that the perspective we occupy when we make such decisions is a perspective disengaged from all of our subjective concerns. The locution raises questions about where we stand, so to speak, when we are making such remarks: ‘Who is the ‘I’ that is making the judgments?’ (Part of what I am doing is providing a response to this question.)

As I mentioned in the last chapter, often what prompts us to ask questions about the practical justification of a particular practical standard is that following that standard has practical costs in terms of the various things we care about. “Is morality normative for me?” A person might be prompted to ask such a question because he cares about his family’s welfare, and morality seems to expect him to sacrifice some of what he might do for them in the name of impartiality. “What’s so important about impartiality,” the concerned father might ask, “that I ought to forego using all of my money on my children’s education in favor of sending some to hunger relief?” It is the father’s commitment to his children’s welfare that provokes the request for practical justification. It is the father’s capacity for reflection that makes it possible for him to ask normative questions about his impulse to follow morality’s demands, but it is his concern for his children’s long-term welfare that motivates him to use that capacity.

Two implications. First, stepping back from an impulse to get some distance from it does not require stepping back to a perspective that is free of any desire whatsoever. Second, there is a serious challenge to the claim that it does. A person who engages in a process of reflection that aims at reflective equilibrium between his various practical commitments, each grounded in subjective concerns, it seems to me, would be a strong candidate for the commendation ‘self-governing’. There had better be stronger reasons to doubt that he is than a mere appeal to “it seems as if there is something “you” that is asking these questions.” In my discussion of
Korsgaard’s own theory, in the fifth chapter, I will consider the claim that autonomy had better not require the stepping back from all of our desires.

A Final Analysis

So, what conclusions are we able to draw from our discussion? What I take myself to have accomplished is this. There are serious objections to the two arguments for the rational impotence thesis; and so, in the very least, these arguments are far from conclusive taken as objections to subjectivism. Also, the indiscriminate subjectivist has a number of ways of interpreting any purported examples of rational impotence in ways consistent with her theory. This provides the indiscriminate theorist with some hope that she is able to accommodate any purported example.

Consider the strengths of the indiscriminate theorist’s response. With respect to any purported real life example of a rationally impotent basic desire, there are legitimate questions that ought to be raised if we are to come to a final, well-considered judgment about whether the purported example is most plausibly interpreted as an authentic judgment of rational impotence. The response, taken as a whole, suggests a stepwise method of evaluation.

First, the desire would need to be basic and to be adequately described. The criticism would need to be rational criticism, not merely an application of the rules constitutive of a substantive perspective, such as morality or prudence. Also, it had better be the basic desire that is coming in for the criticism and not an off limits means. And it is necessary to ask whether the trumping model is adequate to the task. It seems likely that some purported examples should be assimilated to the trumping model; some are rationally potent basic desires that are swamped. I think that the debunking strategies I’ve described on behalf of the indiscriminate subjectivist
theory do deflect some of the force of the anti-subjectivist challenge. And I suspect that there are instances in which a debunking explanation is true: some “discountings” are not judgments of rational impotence but wishful thinkings and performances.

But I don’t think she should rest, assured.

Discriminate Subjectivist Arguments against Indiscriminate Subjectivism

So, why not think that if you are going to go Humean, the indiscriminate theory is better? It has greater theoretical simplicity, the arguments for the rational impotence thesis are not conclusive, and the purported examples of rational impotence admit of analyses consistent with the indiscriminate theory. Why not, then, go with it?

I think that there are several reasons not to. Whatever strengths it has, the indiscriminate response, taken as a whole, is slightly unsatisfying. For one, it turns, in part, on the fact that human psychology is opaque. We often aren’t sure whether the mental states we encounter in the blooming, buzzing confusion of our mental life are basic motivational states. This leaves open the possibility that the purported examples could all be explained away; but it hardly secures this conclusion. It is clearly conceivable that once all these strategies have been considered, a person could still think that a judgment of rational impotence is called for against one of his desires. If nothing else, the advocate of the rational impotence thesis could simply create hypothetical examples and stipulate these things— theorizing has to confront hypothetical examples, too. And so, the indiscriminate theorist has to deal with this question straightforwardly and rely upon the trumping model.
Evidence for the Rational Impotence Thesis

The indiscriminate response to the case of the pacifist with the vengeful desire does not succeed. Let’s stipulate the following about the case. The pacifist is not mistaken about the nature of his vengeful desire. The desire is both basic and sufficiently described; it is, let’s say, the result of an evolutionarily advantageous affect program triggered by features of his present environment. And what he wants, in having the impulse he does, is vengeance *simpliciter*, as it were: to strike back simply because he has been struck at. This impulse is not sensitive to costs, or to moral considerations; it hasn’t yet been processed through the pacifist’s moral ideals. It hasn’t been reformed and reshaped into an apt and morally salutary desire for the re-assertion of his own value as a person. It isn’t even the result of means-end reasoning from a (more) basic desire to win the respect of the other players in the pick up game. The desire to strike back is a desire for “pure” vengeance. It is “raw psychic material.” This pacifist is a human being who has been conditioned by evolutionary history to feel an impulse to strike back when struck at, and he’s been struck at. Is his subsequent vengeful impulse rationally potent for him?

In section II, I imagined a response from the indiscriminate theorist. The theorist attempts to argue that the pacifist’s denunciation of his vengeful impulse is not an instance of rational criticism. It is, she might say, an instance of moral criticism, and moral standards are not, in and of themselves, rational standards. Or it is criticism, more directly, from his pacifistic ideal, which is likely to have moral content, but this ideal isn’t rationally mandated, either.

It is true that morality and pacifism are not universally valid practical standards. But the conclusion that the pacifist isn’t engaged in rational evaluation of his basic desire when he seeks advice from his personal ideal does not follow from this. A standard does not have to be intrinsically rational for it to provide legitimate rational advice to an agent. If an agent cares
about the standard, other things being equal, the standard’s reason claims are practical reasons for him. A personal ideal functions as a normative standard. It is uncontroversial, among (non-minimalist) subjectivists, to think that the personal ideal, if he cares about it, ought to shape his overt behavior: a good pacifist can’t properly strike back. And it also strikes me as uncontroversial to think that the agent’s personal ideal ought to shape how he weighs his practical reasons: a commitment to pacifism renders vengeful impulses, in the very least, very weak. I don’t see any principled reason for thinking a personal ideal can’t also properly shape other aspects of practical deliberation, how he ought to collect his practical reasons.

The indiscriminate theorist’s way of regarding a personal ideal treats the desire that grounds the personal ideal as “inside” the will, a rationally potent basic desire, and the norms that are constitutive of that ideal as “outside” the will. I think, on the other hand, that in caring about the ideal, the norms constitutive of that ideal are incorporated into his will and are valid norms of practical reasoning for him, relevant to rational advisability. If there are any norms of rational impotence, these norms become, in standard cases, volitional norms—collection principles—for him.\(^{129}\) True enough, the pacifist’s ideal is not rationally binding on him *qua* agent. But it is, in my view, rationally binding upon him in virtue of his caring about it. This is an instance in which an agent cares about a very complex state of affairs: he wants to exemplify his pacifistic ideal; he wants both to behave and think as a pacifist would. He wants to “embody” the ideal.

This also presents a different way to think about the purported cases of rationally impotent basic desires that are criticized on moral grounds. Cohon, for instance, thinks that a

\[^{129}\text{The qualification ‘in standard cases’ means to recognize instances in which an agent cares about an ideal but, in the process of pursuing it, comes across norms within it, whether behavioral or volitional, that give him significant pause. Recognition of such norms could conceivably lead him to reject the personal ideal, to reconstrue it by extruding the norms from the ideal, or to set aside his negative reaction to them and go on to treat them as legitimate norms for him.}\]
hateful desire to stick it to a colleague is to be disenfranchised because it is morally objectionable. In my view, she might very well be right about the claim that this desire is rationally impotent for the agent, and she might very well be right that moral standards play a role in rendering it so. What I disagree with is only her suggestion that moral standards are objectively and so universally valid. Attribute to the relevant agent a personal ideal with significant moral content, including norms that forbid her from treating certain basic desires as reason-giving, and I agree that the desire is rationally impotent for her because it is immoral.\textsuperscript{130}

Say that someone objects to my claim that practical ideals are able to have the normative authority to guide an agent’s practical decision-making. Think about an agent who does care about morality. I take it that living the moral life requires a conception of what morality expects. A normative ethical theory is a theory about the content of morality; it aims to tell us what ends (if any) morality expects us to pursue, what rules (if any) it expects us to follow, what (if anything) it expects us to care about, and so on. If an agent cares about morality and accepts—implicitly or explicitly—a normative ethical theory that includes norms that forbid treating some (basic) desires as reason-giving, then his moral commitments are the source norms of rational impotence.

There are, of course, disputes about the real content of morality, and some normative ethical theories do not include norms for practical deliberation, or at least not collection principles. If an agent cares about morality, this entails that some moral standards are a part of her practical point of view, but it doesn’t guarantee that these moral standards include norms of

\textsuperscript{130} A personal ideal has the same basic structure as a substantive normative perspective such as the economic perspective or the moral perspective. It is a practical standard.
rational impotence. But if she cares about moral standards that do include them, then, other things being equal, they are a part of her will.

In summary, the discriminate theory handles personal ideals and other relevantly complex practical standards better than its indiscriminate cousin.

Other Advantages

The discriminate theory has at least two other advantages. The discriminate subjectivist theory presents a unified theory of rationally impotent basic desires and off limits means. Its appeal to personal ideals as the sources of norms that require the categorical rejection of basic desires can be extended to account for the categorical rejection of action-types, too. The person who wants to live up to the conventional marital ideal cannot treat infidelity as a practical option. Even if being unfaithful were a means to a desired end, the behavior is “off limits.” Were the indiscriminate theory to admit the categorical rejection of action-types, the “off limits means,” but deny the categorical rejection of basic desires, her theory would lack a natural unity.

Also, the discriminating theory has an easy way to handle motivational states that arise within the contexts of thought experiments or literary and dramatic experiences.

In the case I considered earlier, an agent experiences what he calls a suicidal “whim,” but it turns out that he does not, in any sense, desire his own death. He is merely imagining that he wants this. Though I think both that this is a real possibility and that the indiscriminate theory is able to handle it with ease, I think there are similar cases that my theory handles much more naturally than hers does.

Consider being lost a book. In the midst of reading, you begin to see the world through the eyes of one of the characters—through his beliefs and values, and in his social and personal
circumstances. You begin to feel righteously indignant when the character suffers an affront. You find yourself wincing when he suffers a blow. You begin to feel the pull, characteristic of desire, of things he wants, things you’ve never felt a desire for before. Your identification with the character is, in the moment of reading, strikingly complete. “He’s occupying my mind, and even my body!,” you joke. And in the midst of all this identifying, you feel the push or pull that constitutes a suicidal desire.

Or, taking the anthropologist’s outlook, you wonder what it would be like to care about something that you have not cared about before. What would it have been like to be a Homeric warrior? What would life have felt like in that social world with its emphasis on nobility? If you begin to get a feel of what it would be like to think and to feel and to desire in the way a Homeric hero might have, what you are experiencing is a piece of imagination. It seems sensible to say that, though in the midst of the imagining I show certain tell-tale signs of having a belief, I don't really believe those things. Likewise about desire. You might, after such reflection, begin to want something you have not wanted before, even something that is objectionable from your practical point of view. But you also might not. The want might stay only within the thought experiment and not have an effect on what you actually care about. You had an interesting experience, but it didn’t change the fundamental way that you care.

A drug counselor might want to know what it is like to be under the influence of an addiction, so that he is better able to know how to treat it.\(^\text{131}\) Not wanting to develop an authentic addiction, with all of its horrific consequences, the counselor imagines having the desire, attempting to see what it would be like to be someone who is consumed by such a drive. In the midst of his imagination, the counselor might begin to experience some of the standard features

\(^{131}\) Frankfurt considers such a case (“Free Will and the Concept of a Person”).
of addictive desire, but the psychological state does not count, in the relevant sense, as a desire. The psychological state generated by this thought experiment is a piece of imagination, not a candidate reason, to be examined by the will.

The indiscriminate theorist might argue in several ways. In this case, the push or pull you feel is embedded within a thought experiment; it is not a part of the ordinary course of your life. The psychological state, the indiscriminate subjectivist might say, is not appropriately described as one of your desires. Hence, this type of psychological state—if it is properly characterized as a motivational state—does not seem to bear the standard connection with having a practical reason—in the real world, at least. Are we to imagine that the Humean theory forces its adherents to give over authority to the motivational states that arise in the context of a thought experiment like this? But how to square this with the general inclusivity of subjectivism? The indiscriminate theorist has, to this point, resisted making appeals to the ordinary pattern of desires and so on to rule out whims and urges and so on. If being a basic desire is sufficient for rational potency, the indiscriminate theory does seem compelled to give it a voice. Mine doesn’t.

In losing yourself in a good book, you might identify with a character who has a very different practical point of view from yourself and begin to experience pushes and pulls that accord with the character’s worldview. These motivational states—though in the gross, literal sense they belong to you—are not really your desires. They don’t fit into the set of pursuits at the heart of your life; they aren’t passions that your reason is enslaved to. Even if they are endorsed by a set of values that the character is rationally committed to, they don’t connect in the relevant way to your goals and ideals. If they violate your standards, you do have grounds to treat them as rationally impotent.
Some Implications

My theory emphasizes the importance of the concept of an agent’s practical point of view, the vantage point from which he ought to make his practical decisions. A loving mother’s commitment to her son’s welfare is a part of her “practical identity.” My contribution, if I am correct, is to attach norms of rational impotence to such an account. Ruling passions are able to “commission” norms of rational impotence to structure practical deliberation.

My proposal also indicates why a rather wide range of purported examples seem plausible. It is not a desire’s duration, its etiology, or its causal power that determines whether it is rationally impotent. It is a matter of the content of her personal ideals. It’s the desire’s violation of the agent’s personal ideals that renders it rationally impotent. But it does follow from my proposal, attractively, that quirks of physiology (such as motivational states caused by brain lesions), bizarre manifestations of the subconscious (such as hostility to parents), and the residue of our past inadvisable behavior (such as motivations caused by addictions), are not guaranteed a voice in our practical deliberation. If they violate a valid norm of rational impotence, they are ruled out. My proposal also suggests that agents who do undergo radical changes in the will akin to conversions do not have to cede authority to old patterns of desire that their new outlook frowns upon.

Does the concept of a norm of rational impotence seem mysterious? I could imagine someone wondering where these norms of rational impotence come from. My view is that norms of rational impotence derive from various sources, none of which are mysterious.

Here’s one source. Kant thinks that what a fully rational being would do is what you and I should do. His behavior is normative for us, who are capable of reason but less than fully rational. In this respect, we might attempt to imagine how he would behave—how he would
think through practical decisions and how he would behave—and mimic him, or at least try to. An agent is to structure his deliberation in the way that the exemplar does. I think some personal ideals do operate in this way: the agent has an exemplar of his ideal in mind and attempts to mimic the paradigm’s (imagined) patterns of feeling, reasoning, and behavior. Think of the “What Would Jesus Do?” bracelets some Christians wear. The bracelet is intended to prompt this type of reflection and subsequent emulation. Also, the loving mother of one of our main examples might wonder what a good mother would be like and take this imagined exemplar’s character and behavior as normative. It doesn’t strike me as implausible to think that real, flesh-and-blood mothers might think the following thought: “a good mother doesn’t treat violent impulses as reasons, and I want to be a good mother.”

I take that many norms of rational impotence, such as the mother’s, derive from social roles. These norms are contingent, and perhaps highly contested. We could imagine a parent confused about just what his role demands of him. A father might wonder, “Should I pay for my children’s college education?” Society isn’t likely to give him uniform advice, and I doubt that the question admits of a single right answer. I could imagine perfectly decent fathers coming to different conclusions. Hence what one of your ideals really requires of you is something you might need, partially, to negotiate yourself. But I find it natural to describe the question the father is asking in terms of the norms of fatherhood: “What, as a father, should I do?” “Is it a rule that a parent ought to pay for college?” Likewise, I take it that an agent attempting to live up to the role he occupies might well bump into questions about how he ought to treat the various impulses and desires that well up in the midst of life. It is natural, similarly, to cast these questions in terms of ‘norms for what to treat as reason-giving’. There’s nothing mysterious here. The sense that there are norms of rational impotence arises in the course of attempting to negotiate a complex human
life in which a person wants to exemplify an ideal but finds in herself impulses that violate these ideals.

My account cuts up the set of all basic desires, in a very intuitive way, into three classes: there are ruling passions, “mere” desires that are rationally potent, and “mere” desires that are not. Many “mere desires” do not come in for deep criticism from the agent’s practical point of view, and there is no problem with treating them as generating practical reasons. When, with respect to what is dear to you, nothing much is at stake, such a mere desire might even generate a final practical reason of rational advice. My desire for ice cream is like that. My basic desire for ice cream is not among my ruling passions; it is simply something that I happen to want. The example of my basic desire for cookies and cream ice cream is likely to count as a desire that is neither a ruling passion nor a rationally impotent desire. It generates reasons for acting, but tends to lose out in competition with basic desires that do have a more privileged status in my practical deliberation. When my child is in danger, such a desire will hardly receive, and hardly deserve, any voice in my practical decision-making, though it won’t be utterly rationally impotent.

More Contested Ground

As Frankfurt construes the case of the unwilling addict, the addict “determines” that his present desire for heroin is rationally impotent, and he subsequently “extrudes” the desire from the context of decision-making. Let’s call a claim that states ‘This basic desire is rationally impotent’ a “judgment of rational impotence.” Frankfurt seems to think that judgments of rational impotence admit of truth and falsity. A realist or a Kantian might latch onto this feature of Frankfurt’s example and ask what the “truth-makers” for these judgments of rational impotence are. For instance, in “The Roots of Reasons,” Cohon claims that practical reasons are
“discovered, not made” and takes the “fact” that we “discount some desires altogether” as evidence that there are realist standards, standards that are independent of our choices, by which we are to judge our (basic) desires, to see whether they really give us practical reasons. We don’t, she says, “construct” these standards, we find them “ready-made.”

Within my proposal, norms of rational impotence, at least many of them, do display a type of independence from personal choice, but I don’t appeal to realism account for this. A judgment of rational impotence is true if the agent is rationally committed to personal ideals that contain norms of rational impotence that rule out the basic desire in question. Even if the contents of a personal ideal are socially constructed, they might not have been constructed by the particular person who discovers them through reflection upon his ideal. He might find them ready-made.

The second claim is that an act of will plays a role, some role or other, in disenfranchising a basic desire. In Frankfurt’s example, the unwilling addict “extrudes” his craving for heroin by an act of will. I think an analogy to a judge who interprets the law and subsequently issues a ruling is helpful. Akin to a judge who applies legal standards to particular cases and makes “official” declarations, the will is able to declare, on the basis of norms of rational impotence, that a particular basic desire is rationally impotent. And once a determination is made that the basic desire is to be disenfranchised, the will’s extrusion is on analogy to a judge’s official verdict. The judge makes “official” what the rules require; the act of extrusion makes “official” what the valid norms of rational impotence require.132 I will give acts of will, ...

---

132 As evidence for these claims, I think it is possible to conceive of an agent who gets the rules wrong, or applies them mistakenly, which is enough to ground talk of truth and falsity. Also, it is possible to conceive of an agent who recognizes the rules and their implications, but simply fails to extrude the basic desire in accordance with his recognition of what the relevant norms of rational impotence call for.
then, a *subsidiary* role in disenfranchising basic desires. These acts of will are done in service of the agent’s ruling passions.

My remarks are pre-emptive. I don’t mean to suggest that realists and Kantians are likely to find an appeal to the intuition that judgments of rational impotence admit of truth and falsity or an appeal to the intuition that the will has the capacity to extrude a basic desire to be the most profound evidence for realism and Kantianism. But they are intuitions that a theory about the phenomenon Frankfurt is gesturing at ought to account for. My subjectivist proposal, I am pointing out, has the resources to account for them.

Another Proposal, Briefly Stated

Question: does the will ever have the prerogative—the normative authority—to rule out a basic desire independent of norms of rational impotence that are “already” a part of a personal ideal? I’m inclined to think so. Bratman talks of “plans,” and he construes practical reasoning as consisting largely in the adoption, filling in, and reconsideration of plans (*Intentions, Plans, and Practical Reason*). Plans are typically incomplete—or “partial”—and the stable intention to follow a plan often requires an agent to adopt “sub-plans” at some future date, to execute the larger plan successfully. A person who plans to go to the airport in Madrid in a few weeks might leave it until the day of his flight, when he has spent more time in Spain and knows more about the transportation system there, to decide upon a sub-plan for how to go about getting there. Plans tend to be stable and flexible, they can also be partial, in need of further specificiation.

Given that many personal ideals are, to one degree or another, vague or indeterminate, an agent who adopts a personal ideal might have occasion to fill in details that are not already present in the ideal. I don’t see any principled reason to think an agent could not adopt a personal
ideal with the same type of “partial” structure. He might resolve to become a good, or a better, father. He is likely to have some thoughts about what this ideal expects, but in all likelihood he doesn’t have it all worked out. In the process of specifying the content of his ideal, he might adopt a norm of rational impotence as a way of (further) specifying the content of his own ideal.

In the introductory chapter, I said that I intend to explain “one way” that a subjectivist is able to account for the rational impotence thesis. This is another way. But I won’t attempt a fuller discussion. I will consider it a sufficient defense to reveal that norms of rational impotence with subjectivist groundings do exist and do have the relevant normative authority. The will, in the examples I intend to defend, is operating in accordance with norms of rational impotence grounded in a ruling passion.\footnote{There might also be instances in which the will engages in a radical act of self-creation. Confronted by a desire that seems to him to be objectionable, he adopts, say, a personal ideal that makes sense of his particular revulsion to the desire. In this case, he incorporates into his practical point of view a personal ideal.}
CHAPTER 4

DOES A KANTIAN THEORY BEST CAPTURE THE RATIONAL IMPOTENCE THESIS?

I affirm the rational impotence thesis: there are (logically satisfiable) basic desires that do not generate practical reasons. My central question: how are rationally impotent basic desires rendered rationally impotent? What makes them so?

You might think Kant is well-equipped to answer this question. It is central to Kant’s theory of practical rationality that practical reason itself stands in judgment of any desire and the claims it makes, and it disenfranchises some of them—for example, any immoral inclination. Kant, it seems, generates a theory of the will that is geared up to make sense of the idea that the will has the capacity and the practical authority to, among other things, issue “inexorable demands without making any promise to the inclinations” (405). The will “is” practical reason, and it does not consist of any of the agent’s desires. Practical reason is a set of rules of rational willing.

To say the least, the differences between my subjectivist proposal and Kant’s general theory run both deep and wide. In my subjectivist theory, practical reasons—all of them—derive, ultimately, from among the agent’s subjective motivational states. There is a standard connection between having a basic desire and having a practical reason to take the means to its satisfaction. Though I think that this link is sometimes severed, by a norm of rational impotence that operates
as a defeater, I think that practical reasons are standardly and relatively directly communicated from desired end to effective means, by way of an instrumental principle.

Kant thinks, on the other hand, that a satisfiable basic desire, in and of itself, does not generate a practical reason. As Korsgaard puts it, a desire presents a “problem,” and “not a reason.” In the strict sense, every basic desire, in and of itself, is rationally impotent. At the same time, it’s important to recognize that, in Kant’s view, basic desires do have a significant role to play in the generation of practical reasons. There is a connection between an agent’s desired ends and his (final) practical reasons. Rational action has an aim in view, and our aims are defined in terms of our desired ends. But in Kant’s theory, the route from having a basic desire to having a practical reason is less direct; an inclination to act is (for lack of a better phrase) “associated with” a principle of action—a “maxim”—and there is a test of practical reason, the Categorical Imperative, for the rational permissibility of any maxim, for whether the maxim is fit to be acted upon. If a maxim fails the test of the Categorical Imperative, it is not rationally permissible for an agent to act upon it—even if doing so would satisfy a desire that is very dear to the agent’s heart. If the Categorical Imperative condemns a maxim, we might also see it, by extension, as condemning the inclination to act associated with it. Kant consistently construes inclinations to act as making claims upon the will: ‘Do this. Do this’. If the inclination’s associated maxim fails the test of practical reason, the inclination does not play a role—even indirectly, through its maxim—in generating a practical reason for the agent.\textsuperscript{134}

In this way, without doing violence to Kant’s language and conceptual scheme, we could construe the Categorical Imperative as a “kind of test of normativity” for desires. If a desire’s claim does pass the relevant test of reflection, then desire deserves to be heard by the will, and its

\textsuperscript{134} An appeal to the basic desire would explain the agent’s behavior, but not justify it in the relevant sense.
claim is to be weighed against other legitimate claims. In this way, Kant theory does aim to respect the differences between people with respect to goals and personal projects: what agents have a practical reason to do will often differ and in part because they bring different desires to the test of practical reason. The heart’s desire has to be put to a test, but it does get to supply the input for the test.

My aim in the following three chapters is to challenge both Kant’s theory of practical rationality and a contemporary attempt to refurbish it, Christine Korsgaard’s. There are two reasons why I am taking up this project, one destructive, one constructive. First, the destructive task. I accept a subjectivist theory of practical reasons. Thus, I have an interest in undermining my subjectivist theory’s competitors, and Kantianism is a major competitor. Sometimes a good defense consists of a strong offense. I think my theoretically modest proposal for how to account for the rational impotence thesis compares favorably with Kant’s and Korsgaard’s theoretically more robust proposals.

Second, the constructive task. I think a refined subjectivist theory of practical reasons is able to capture several insights—related to the character of the will and its authority to stand in judgment of some desires—that are often attributed to Kant and to Kantians. I hope to steal away these insights for my own theory. I propose that a refined subjectivism is able to capture them without suffering from (what I take to be) compelling objections against Kant and Korsgaard. I want to mine out the gold, even as I undermine the theory.

In particular, I want to steal away (something very similar to) Korsgaard’s concept of a normative self-conception, a “practical identity,” which she employs in her defense of a refurbished Kant, and make it a centerpiece of my refined subjectivist theory. As I see it, Korsgaard is correct to think a practical identity—I have spoken, instead, in terms of “personal
ideals”—provides an agent with a practical point of view from which to critically evaluate, among other things, his own inclinations and to judge some rationally impotent. I don’t think, though, that the concept of a practical identity, once fully understood, fits into a Kantian conceptual scheme. Korsgaard fails—and because of her Kantian commitments—to give an adequate account of how a practical identity might come to have legitimate practical authority for an agent. My proposal does better. A personal ideal or a practical identity is valid for an agent only if it is grounded in the agent’s ruling passions. Inclinations are rendered rationally impotent only by standards that are themselves grounded in subjective desires. Hence, the idea of a practical identity coheres with subjectivism, not Kantianism.

The Program

The three chapters take the following course. In chapter 4, I will describe a general normative question that Kant, Korsgaard, and my subjectivism answer differently, and I will give at least an initial and very general description of how Kant and Korsgaard interpret and answer the question. Things are going to get complicated—with a revolutionary thinker such as Kant, how can they not? So it is useful to give a general sketch of their views early on.

In chapter 5, I am going to describe Kant’s theory in greater detail, along with several common objections to it. These objections will provide my primary reasons for rejecting Kant’s proposal for how desires are properly disenfranchised. Also, since Korsgaard’s own theory is intended to be Kant’s theory, adjusted to overcome these objections, a discussion of these

---

135 One advantage that Korsgaard’s theory has over Kant’s, in my view, is that it does not tie norms of rational impotence as tightly to moral norms as Kant does. Korsgaard thinks that norms for rational impotence reside within an agent’s self-conceptions, and many self-conceptions are not in and of themselves moral standards. This strikes me as correct, but a significant deviation from Kant.
objections prepares us for the chapter about Korsgaard. In chapter six, I will describe Korsgaard’s refurbished Kantian theory and the central place that the concept of practical identity has within it. I am also going to argue against Korsgaard’s theory on several grounds. In particular, I am going to argue that Korsgaard’s theory is equally susceptible to several of the common objections to Kant that she intends for it to overcome, and—most importantly for my purposes in the dissertation—that being governed by a practical identity is incompatible with Kant’s conception of freedom and heteronomy, which are aspects of Kant’s theory she doesn’t—and can’t—give up. My ultimate goal is to argue, in the next chapter, that the concept of a practical identity better fits within a non-Kantian, refined, subjectivist theory of practical reasons.

What’s the Question?

Kant thinks that all of us experience, some of us more often and more intensely than others, an internal struggle between the Moral Law and our inclinations (Foundations, 439). Our desires are often felt to be, in Kant’s words, a “powerful counterpoise” to the demands made by morality (405). The eye of reflection might notice this conflict and lead us to wonder why we ought to favor morality, or even why we ought to give it and its claims any hearing whatsoever. Reflection is able to call out morality to justify the demands it makes upon us—not to mention, its right to make us feel guilty when we break its laws.

Take the following scenario, to illustrate. A good job (or grade, or love interest) is your heart’s desire. Told not to cheat when you could get better hours and a corner office through

136 Quotations of Kant are from Lewis White Beck’s translation of the Groundwork. Citations refer to the standard Academy pagination that most translations, including Beck’s, provide.

137 Listen to Kant, beautifully translated by Beck, “Man feels in himself a powerful counterpoise against all commands of duty which reason presents to him as so deserving of respect; this counterpoise is his needs and inclinations, the complete satisfaction of which he sums up under the name of happiness” (405).
deceit, you ask yourself ‘Why forego a better job for honesty?’ Told that honesty is a moral duty, you wonder why you ought to care about that. In this scenario, the voice of conscience—yours, let’s imagine—is speaking to you, telling you what (not) to do, and you’re wondering whether, and why, you ought to lend an ear to morality’s demands. “Do moral reasons really deserve,” you’re asking yourself, “a voice, any voice whatsoever, in my decision-making?”

Korsgaard calls this request for practical justification of morality “the normative question” (Sources, 7-18). And she presumes that you and I, her readers, identify with the struggle that commonly leads to the normative question; as she puts it, you see this struggle as recognizably “yours.” As a self-conscious creature, presumably you have experienced this internal struggle and have wondered, from time to time, whether and why you ought to listen to your conscience.

Korsgaard follows Kant’s habit of personifying our mental states and mental powers. Kant and Korsgaard accept Plato’s political analogy for the soul. Korsgaard consistently speaks as though the mind has a variety of inhabitants, interacting with each other—sometimes in conjunction, sometimes in conflict—within the economy of the soul. Various constituents come

---

138 Or, in Korsgaard’s words, what “right” does the Moral Law have “to give laws to us?” (9); or “what justifies the claims that morality makes on us?” (9-10, emphasis hers).

139 See Korsgaard’s remarks about the normative question as a “first-person that arises for the moral agent who actually has to do what morality says” (16).

140 For a discussion of Kant’s tendency to hypostatize our various powers, see Roger Sullivan (Immanuel Kant’s Moral Theory, 193). Kant oftens says things such as “empirical practical reason,” that is, prudential reason, ‘does this’, and “pure practical reason” ‘does that’.
to the negotiating table to speak their minds. I’ll follow Korsgaard’s anthropomorphizing habit, for the time being.

In Korsgaard’s analysis, there are three participants in the internal dialogue prompted by the dispute: the voice of conscience, of the heart’s desire, and also of reflection itself. To understand what “the voice of reflection” is, think of the Cartesian ego, the res cogitans of Descartes’ Meditations. You and I have the capacity for reflection: the ability to step back from our mental states and to ask questions about them. This capacity to take our own mental states as the object of reflection is the defining feature of self-consciousness, a part of the “structure” of the reflective mind (92-93, 96). And in Korsgaard’s view, the “reflecting self”—as “ephemeral” as it is (228)—is a third party to the dispute between conscience and the heart. Taking its characteristic step back, it is able to take the conflict between the other two parties into its gaze and, as an impartial judge might, ask for practical justification from either or both disputants.

This brings us to a point crucial to grasping what question, precisely, Korsgaard is talking about. You are asking the normative question if, and only if, you are making a request for practical justification of morality from the perspective of reflection itself. Reflection is able to call out the heart, as well, to make it justify itself and its demands. As I’ll discuss in a moment, Korsgaard thinks this is a natural and legitimate question for reflection to ask. The heart makes commands, and we can sensibly wonder whether they are normative. (This is the thesis that even the indiscriminate subjectivist should grant, I argued in chapter 2.) But Korsgaard would insist that you’re asking what she calls “the” normative question only if you are calling upon the Moral

---

141 Kant and Korsgaard also share Plato’s sense that in many human beings this interaction does not make for a well-ordered community, largely due to the influence of unruly appetites. Kant and Korsgaard also share Plato’s belief that a good human life displays a harmony in the soul, with reason in the ascendancy, “ruling.” But there are important differences in their conceptions of the dictates of reason and of justice.
Law to make its case, and only if you are asking this question from the perspective of pure, primitive reflection.

It’s hard not to read into the scenario, and so into the normative question itself, a self-serving tone. You really, really want that promotion. So, we could easily interpret the scenario as an instance in which the dear self is doing the asking. It has become resentful of morality and all of its constraints. (“Morality,” it grumbles, “so often says ‘No. No. No.’”) Petulantly, it enlists the capacity for reflection to serve its own ends. Its strategy is to weaken the grip of conscience by asking difficult skeptical questions of it. In short, the dear self aims to win the conflict with conscience by strapping it with a huge task.

Korsgaard would reject this interpretation, on several grounds. If the scenario is going to count as an illustration of the normative question, it has to be read in a different way. Two remarks help to clarify, the second more to the heart of the matter than the first. First, Korsgaard conceives of the normative question as a request for practical justification, all things considered, and self-interest is only one consideration, among many. The human heart consists of—at least for many people—more than the fat, relentless ego, and so the demands of morality can come into conflict with a wide range of personal motives. Maybe the reason you intensely desire better hours is to spend more time with your children, to benefit them. And a person who feels natural sympathy for the long-suffering family of a victim of a serious crime might wonder why we ought to extend the moral right of due process to someone as unsympathetic as the

---

142 It takes time, and smarts, to answer the question ‘What right do you have to tell me what to do?’ That’s why this question can be an effective tool in political contexts.

143 Korsgaard thinks that Kant tends to oversimplify when he construes the confrontation between the Moral Law and the heart as a conflict between the Moral Law and the dear self. She means to remedy this oversimplification in her theory.
victim’s unrepentant and undeniably guilty perpetrator. Pertinent to my subjectivist proposal, the demands of conscience could conceivably conflict with the demands of a cherished personal ideal. For this reason, Korsgaard does not conceive of the normative question merely as a prudential question: ‘Does honesty pay?’ or, more generally, ‘Is moral virtue in my best interest?’ Whenever a person’s heart is bigger than his ego, reflection could compel him to ask ‘Why should I obey conscience if and when it conflicts with my heart’s desire?’

But more to the heart of the matter, the normative question is not a question asked by any voice from within the human heart. The normative question, as Korsgaard sees it, might be prompted by your personal inclinations. But properly understood, the normative question is not asked on behalf of them. It is asked by pure, primitive reflection, the capacity to step back and ask questions. If the reflecting self must be described as driven by a motive, it is a natural desire to know. But the relationship between reflection and this motive is special. It is simply a part of reflective consciousness, “the structure of the reflective mind” (93), that we ask why?, and the question why?, asked by pure reflection, is simply an instance of intellectual curiosity. The reflecting self does not have an agenda, as the dear self or, more broadly, the heart. (It’s innocent in the way Socrates, as described by Plato, is.)

If this idea of pure, primitive reflection seems strange or problematic, there is an analogy to how reflection is able to operate upon our beliefs. You might be prompted to reconsider a belief by a personal motive. But once embarked, you might find yourself wondering whether your belief is true, independent of your motive. Likewise, with respect to what Korsgaard calls the normative question, reflection is directing its attention at the Moral Law and asking whether it

---

144 Korsgaard considers the case of a lawyer whose moral duty to work on behalf of a worthless client does not serve her more heartfelt end of increasing social utility (86-88).
is normative. Even if a personal motive played a role initially in bringing you to ask for practical justification, once you have noticed the voice of morality, the normative question itself is available. That is, according to Korsgaard, “simply how reflection works,” it is at the very heart of “the structure of the reflective mind” (93).

Back to which features of the scenario are salient. In the scenario, it seems that the prospect of passing up the new job causes you to feel intense despair and personal anxiety. But these feelings are not essential to the normative question. Korsgaard herself and her commentators, wanting you to see how important to human life the normative question is, tend to choose illustrations (as have I) that pit conscience against a particularly intense desire of the heart. But Korsgaard thinks the normative question is “pervasive”; it is present whenever morality tells, or even encourages, us (not) to do something. But the normative question is not always “dramatic” (13); not everything it tells us to do forces us to give up something we really want. So, there doesn’t need to be a struggle for us to ask for practical justification of morality. Even if you find yourself wanting to do what it says, Korsgaard thinks we could still ask whether its demands deserve a voice.

That said, Korsgaard does think that morality has a “particular kind of importance” to us, her readers (11). Most of us, by the time that we reach the age of reflection have long since internalized some moral norms. To one degree or another, moral education has taken in us. And morality exerts for many people considerable psychological influence; Korsgaard talks of how some people have been willing to die rather than violate moral standards (11). Korsgaard calls these “the practical and psychological effects” of moral ideas, and she thinks the normative question generally is an important question at least in part because so many of us are affected in
deep and important ways by conscience. But Korsgaard recognizes that even if we do treat the demands of conscience as valid, we can still ask whether the influence it wields is warranted.

Korsgaard’s Rejection of Sophisticated Subjectivism

Korsgaard thinks that reflection upon morality’s demands on behalf of the heart is a legitimate and important task. A thoughtful person will consider whether and to what degree a life of obedience to the Moral Law accords with his various personal motives, the desires of his heart. If you want the best for yourself, if you love your child, if you happen to want a troubled family to come to a resolution or for humankind to flourish, you can sensibly wonder “Would these interests be served by my obedience to the Moral Law?” Morality might well ask you to do something, in a particular case, that doesn’t serve these desired ends. In my terminology, this is the question whether obedience to the moral law is rationally advisable for you. Korsgaard affirms that part of what it is to be a rational and personally autonomous agent is to reflect upon what you care about and how your various cherished goals and projects fit together. My own general theory claims that practical reasons, in the sense relevant to rational advisability and the rational appraisal of action, are ultimately grounded in contingent, subjective motivational states: basic desires. I’ve also claimed the norms of a personal ideal are valid for an agent if the ideal itself is grounded in one of the agent’s ruling passions and is consistent with other elements of his practical point of view.

Korsgaard calls the attempt to justify a practical commitment to the moral life by this method “reflective endorsement” and attributes it to David Hume, John Stuart Mill, and Bernard Williams (49-89). Korsgaard interprets Hume as a modest defender of the moral life. In her interpretation, Hume claims that human beings are endowed by nature with two significant
motives: self-interest and moral sentiments, such as sympathy, which are at least commonly fostered by society. Through a process of socialization, human beings commonly become persons for whom a moral life is rationally advisable, as well as a source of personal esteem. When we reflect upon the moral commitments nature and society impart to us, “we come to think that we would not be better off without them” (83-84). If so, our moral commitments receive an endorsement from reflection upon our heart’s desire, and we do not have a serious or compelling practical reason, in terms of these motives, to try to become a different kind of person, one for whom conscience deserves little or no hearing before the will. My impression is that Korsgaard thinks that Hume’s reflective endorsement of the moral life has a degree of plausibility. (There is a certain optimism both in Korsgaard and in Hume.) Korsgaard seems to accept that the human heart and the Moral Law are not condemned by their very nature to conflict, at least not deeply and insistently.145

But Korsgaard thinks, as Kant would, that Hume’s method of reflective endorsement, even if it turns out to have favorable implications for the moral life, cannot issue in the proper type of endorsement of the Moral Law. For one, any such endorsement is conditional—‘hypothetical’ in Kant’s language. Even if it turns out that a practical commitment to morality is rationally advisable for you, given your personal motives—or even for most of us human beings, given ours—this endorsement will not apply to every conceivable rational agent, only to those with a sufficiently similar psychological profile and social context. We could easily enough

145 Korsgaard does think that there will be instances in which the Moral Law forbids a course of action that our moral dispositions favor. She thinks this is a serious problem for Hume; it introduces a conflict within the practical deliberation of a person who grounds his commitment to the moral life in his moral sentiments (86-88). For a response on behalf of Hume, see Blackburn (Ruling Passions, 222-223).
conceive of an agent who does not share our concerns, for whom obedience to the Moral Law is not instrumentally rational.

Also, any such endorsement fails to accord the Moral Law its proper grounding. In Kant’s and Korsgaard’s views, the Moral Law, however personally satisfying a moral life might turn out to be, has its grounding in something much deeper than the human heart: in our identity as rational beings. In asking any questions, such as the normative question, what follows is the “inescapable” fact that you and I are creatures with the capacity for reflection. At least in the moments that you are asking questions, you are a thing that reflects: a creature with a reflective mind. Korsgaard has her own version of Descartes’ “Cogito.”

So, Korsgaard treats the reflective endorsement method, reflection upon the desires of the heart from the perspective of the desires of the heart, as an important activity in the life of a rational and personally autonomous person. My proposal for the role of personal ideals, since it is aligned with this type of reflective endorsement, could also be incorporated into this account. That said, Korsgaard rejects Hume-style reflective endorsement, taken as a theory of the ultimate source of practical reasons, or in her words as a theory of “normativity” (Sources, 85-89).146

Korsgaard and Kant think the Moral Law must have a grounding in something deeper than inclination, in “the unconditioned.” According to Kant and Korsgaard, this something is your faculty of reason; it is having this faculty that makes you what you are. Kant and Korsgaard think that the faculty of reason—theoretical and practical—is a part of our very identity as human beings.

---

146 I think Korsgaard’s remarks are misleading. In the subjectivist’s view, it is not the “endorsement that does the work” of making, say, morality normative for an agent; it is the fact, if it is a fact, that following moral norms is rationally advisable in terms of the agent’s own subjective concerns. The caring, along with the facts about its satisfaction, “does the work.”
beings, and practical reason provides rules for rational willing that necessitate respect for the Moral Law from any rational agent.

This places a significant theoretical burden upon Kant and anyone who follows him. Does the voice of practical reason itself, given that it is thoroughly formal, have the ability to provide substantive practical advice about what (not) to do and what (not) to treat as reason-giving? In the introductory chapter, I said that subjectivists who accept the rational impotence thesis have to make sense of how the will, defined in terms of an agent’s ruling passions, is able to rule out (other) desires. But Kant has his own challenge—to indicate how practical reason, construed as a set of thoroughly formal rules of rational willing, are able to issue in substantive principles for the collection and comparison of practical reasons.

What Does Korsgaard Intend to Accomplish?

Korsgaard herself asks the normative question from at least two motives. The first is the motive already mentioned: intellectual curiosity, or a natural desire to know. Philosophy begins in wonder. Korsgaard wants to know what the right answer is. Her book is her search, written down. Second, Korsgaard desires to advocate for the Moral Law. Similar to Kant, she thinks she has come by the true answer to the normative question, and it is ‘yes, and for everyone’. Along with Kant, she accepts what we might call the ‘universal authority thesis’: moral reasons are practical reasons for every rational agent as such. The Moral Law really does deserve a voice in our practical deliberation and has the practical authority to obligate us—to tell us, among other things, what (not) to do, and which desires (not) to treat as reason-giving. (Also, what to feel:
guilt if you’ve done wrong; due pride if you’ve done right.) Her book consists of her argument for this thesis.\textsuperscript{147}

Korsgaard seems to have a third motive: respect for her own humanity. (This is a more important point than it might initially seem.) Korsgaard speaks as though we are under an obligation—an ‘ought’ of some variety; practical rationality, I suppose—to search for answers to whatever question reflection happens to raise. At the end of her argument, Korsgaard will claim that a person is to treat her own capacity for reflection, which grounds her identity as a rational creature, as a “normative part of her identity,” with the authority to give her laws. One of the laws this identity gives her is to “push reflection as far as it will go.” If a person ignores the call for an answer from her reflecting self, she is, Korsgaard thinks, denying the call of the most fundamental and indispensable aspect of her self. Korsgaard doesn’t appeal to the fact of reason, and she will weaken Kant’s Categorical Imperative considerably, but she does not think the reflecting self as “thin and wispy” as it might seem. The reflecting self is able, in her view, to issue “inexorable commands”—to seek out answers to the questions it asks. So, it is not merely a natural curiosity, an itch for knowledge, that leads Korsgaard to seek out an answer to the

\textsuperscript{147}Korsgaard also wants to put the argument to an apologetical aim. She would like to persuade people unto (greater) moral commitment. Philosophical reflection, Kant thinks, is able to increase moral resolve. Listen to Kant, who hearkens back to the story of the Fall in the Garden of Eden, “Innocence is indeed a glorious thing, but, on the other hand, it is very sad that it cannot well maintain itself, being easily led astray. For this reason, even wisdom—which consists more in acting than in knowing—needs science, not to learn from it but to secure admission and permanence to its precepts” (404-405).
Korsgaard agrees with Kant about this, as well. She understands that we, and people we know, lose moral resolve. The advocate of morality might want to ask the normative question, and learn Korsgaard’s argument, to come by something incisive to say to himself when he needs a little help seeing the validity of the Moral Law; or to his children, to advance and deepen their moral education; or to a good friend who has long been committed to moral ends but has begun to lose heart; or even to the skeptic, for the purposes of pure apologetics. The argument serves as a bulwark against our passions on those occasions when they clamor especially loudly and we get anxious that a life of moral duty isn’t worth it, maybe because it is asking us to sacrifice something especially dear to us. And when the dear self, advocating its agenda, asks its big, unwieldy skeptical questions, it is useful for you to have the arguments handy.
normative question; she’s obeying a *command* made by an identity she regards as inescapable, her identity as a creature with the capacity for reflection. Her own argument ultimately depends upon these thoughts. But we’ll have to set this issue aside until near the end of her argument, in the sixth chapter.

A Few Questions

Is the universal authority thesis true? Does the will, as practical reason, have the practical authority to disenfranchise immoral inclinations? Kant’s and Korsgaard’s arguments are ambitious. They think they are able to catch the agent who has misbehaved morally in an inconsistency in the will, and in a type of inconsistency that the misbehaving agent cannot shrug off—at least not without severe costs, the denial of his own rationality and his own self. If Kant were correct, this would be a great boon for moralists. And it would also mean that my subjectivism is mistaken: there are extra-subjectivist constraints upon practical reasons, and practical reason itself does disenfranchise a class of basic desires on the grounds that they encourage immoral behavior. Let’s see if Kant or Korsgaard can make good.

A Few Objections Set Aside, for the Time Being

I want to take a moment to raise a few objections that might have arisen already in the reader’s mind. But I intend to forestall them. Not because I disagree, but because I want us to take the time to work through the Kantian arguments rather than to reject them at the beginning.

Someone might object, up front, to Kant’s and Korsgaard’s stark “counterpoise.” (I would.) Is, as Kant suggests, the Moral Law itself thoroughly dissociated from all of our
Is whatever validity it has for various agents truly independent of their subjective motivational states? Let’s watch Kant at work and see how he attempts to ground the Moral Law in something other than the passions. Let’s see how far his theory is able to take us.

Second, it doesn’t seem that everyone confronts the normative question. First off, the question seems historically and culturally determined; second, once articulated, it seems highly theoretical. As for the first objection, did people from ages past ask the normative question? Did Achilles? Williams, to name one person, would say ‘no’. Williams claims that the Moral Law, as understood by Kant, is a modern idea, an artifact of the Enlightenment, with roots in Christian and Platonist thinking (*Ethics and the Limits*). Kant thinks—incorrectly, Williams would claim—that “reason presents to us” as so “deserving of respect” the Moral Law. It seems that there are cultures whose members did not have a conception of the Moral Law; so they could hardly have asked questions about it. Sullivan, a consistently charitable contemporary commentator, is even inclined to be critical of Kant on this point. He accuses Kant of a failure to critically evaluate the cultural influence of his relatively uniform Protestant Christian climate and the philosophical influence of the natural law tradition, with its tendency to assume that the existence of a relatively thick and uniform human nature (*Immanuel Kant’s Moral Theory*).

If we were to advance this objection against Kant, we would need to be careful about what we mean by ‘Moral Law’, how much content we build into it. Perhaps past people, Kant could say, shared a sense of some but not all of what Kant considers the Moral Law. Though Achilles did not likely have an apprehension of the liberal values of enlightenment morality, he did exist in a society that had common rules—against cowardice, dishonesty, and so on—that

---

148 There are times when Kant indicates that he recognizes that his construal of the passions is overly simplistic. But he is committed in a deep way to distinction between the will and the passions. He himself says that he “robs the will” of the passions.
could plausibly construed as moral rules. This objection, by the way, prompts Korsgaard to
disagree with Kant, and I want to set aside the objection until we have described Kant and
Korsgaard with enough detail to understand how their views differ.

The second objection is that the normative question is so theoretical that it is not
plausible to think that everyone confronts it. An advocate of this objection might concede to
Korsgaard that most western, post-enlightenment people do wonder, at least from time to time,
whether it is necessary to follow moral norms, for instance, when they conflict with the most
cherished desires of the heart. But Korsgaard’s normative question, fully articulated, is a
theoretical question about the ultimate source of practical reasons. How many people ask for
practical justification of morality out of pure intellectual curiosity, motivated by an impartial
desire to know the truth, and not merely for the sake of a personal motive? I’m not so sure
everyone asks sophisticated, theoretical questions. And if Korsgaard were to build into the
normative question the pursuit of “the unconditioned,” how many people share this project? I’m
inclined to think many people haven’t thought to seek out the unconditioned, and there are
people, such as Hume and other thoughtful subjectivists, who self-consciously reject this type of
pursuit, for theoretical reasons. They confront the normative question in the weak sense that they
consider the practical justification of morality, but they do not regard the search for the
unconditioned as “theirs.”

These objections loom. We’ll confront them, later. I will claim that neither Kant nor
Korsgaard is able to give satisfactory responses to them. Korsgaard’s theory attempts to, but it
forces her to deviate from several Kantian assumptions she intends to affirm.
The Normative Question and the Search of the Ultimate Source of Practical Reasons

The question that Korsgaard calls “the” normative question is a particular version of a general type of normative question. Morality is not the only practical standard about which we are able to ask for practical justification. There are many, and it is possible to make this request of any practical standard that dispenses practical advice. For example, prudence and etiquette are also practical standards, and it is possible for a person to ask whether he ought to treat their practical guidance—the considerations and rules they advance—as practical reasons and practical norms for him.

For that matter, practical standards aren’t the only possible source of practical reasons. A person also might purport to be a source of practical reasons. A political authority, such as a monarch, might claim to be the source of practical reasons for his subjects, or a parent for her children. (Hear: “You must obey me!” with the emphasis upon the ‘me’.) A theological voluntarist about practical reasons would think that God’s will is a source, an ultimate source, of practical reasons. A social contractarian about practical reasons would think that the will of a particular (real or hypothetical) legislative body is an ultimate source of valid rules for the members of a society. Also, a desire might be taken to be a source of a practical reason. (By subjectivists, a satisfiable basic desire is taken to be, in standard cases, an ultimate source.) Speaking anthropomorphically, we might say, as Kant does, that an impulse or a desire often “presents itself” as a source of a practical reason, and the subject of the desire is able to ask whether the desire “really” does give him a practical reason to do as it bids. (The question whether there are rationally impotent desires is, I hope, a coherent and interesting question—it’s the central question of this project.) In summary, an agent is able to ask of anything that purports to be, or might be taken to be, a source of practical reasons for him whether it really is.
All of us are implicated in treating some considerations as practical reasons. Human agents, even when they have not adopted practical standards in a particularly self-conscious way, often display a type of practical deference. Our practical deliberation must begin somewhere. Just as there are beliefs that we reason from and not to, there are considerations that we deliberate from and not to. We, as it were, “treat” these considerations as basic and practical-reason-giving. This type of deference is often paid to desires, in cases in which ‘Because I want to’ is the final justification given, but I don’t think that desire is its only object. Such deference could be shown to a rule, to political leaders, to our parents, or to the will of a deity. In some cases, it is the product of enculturation. A person, through his upbringing, has acquired the habit of treating what his parents tell him to do (or its opposite) as practical-reason-giving for him. There are cases in which the child’s deference is motivated by a desire, perhaps the desire not to be punished. But there are also, I think, instances in which these practices exemplify a kind of volitional inertia.\textsuperscript{149} Ask ‘Why do you follow your parents’ rules?’, if “I simply do” is the true answer, he exhibits the kind of volitional inertia I am gesturing at. The practical deference to their will is a part of the foundations of his will. His practical reasoning begins from this, and there is no reasoning to it. (This deference has causes, of course, but not supporting reasons). Similarly, there will be some people who obey what they think God has demanded who do so out of love or out of fear, or out of the idea ‘He deserves it’. But perhaps others are governed by

\textsuperscript{149} For a similar idea, see Frankfurt (“On Caring,” 162, and “Taking Ourselves Seriously,” 3-7).
volitional inertia; they were taught as children, the practice got entrenched, and there does not remain any further justification.\textsuperscript{150}

This type of practical deference, once recognized, simply leads to the question whether and when this type of volitional practice—treating as practical-reason-giving—is practically justified, all things considered. In short, the most general form of the normative question is the question, of anything that purports to be, or is treated as, a source of practical reasons for an agent, whether it really is a source of practical reasons. Or more broadly, of anything that issues practical advice—about how to feel, what character traits to foster, what (not) to treat as reason-giving—whether its practical advice is valid. What the various normative questions about various purported sources of practical advice give rise to is a general question: What makes for a legitimate source of practical advice? This general normative question naturally leads us to seek out—as the title of Korsgaard’s book suggests\textsuperscript{151}—an account of the ultimate source, or sources, of practical reasons. It leads us to seek out a theory, or to put it stiltedly, the “practical-reason-making features” of (legitimate) practical reasons: ‘What \textit{makes} a practical reason (for a person, P) a practical reason (for P)?’ In more colloquial terms, where do practical reasons come from? And how do they come to have whatever level of practical authority they possess?

\textsuperscript{150} Perhaps you’ve had the experience of recognizing, for the first time, an assumption that you’ve made all along, and which has shaped your thinking profoundly. You simply hadn’t been self-conscious of the idea that has previously exerted such influence. The same experience can happen with respect to volition states.

\textsuperscript{151} It seems to me that there are types of value or normativity that are not conceptually connected to practical reasons in the sense being discussed here. A practical reason, as defined here, is a consideration that deserves a voice in an agent’s practical deliberation. A brilliant painting might have intrinsic value, and I might acknowledge that it does, thereby judging it to have intrinsic value, without there being any practical reasons (of rational advice or of rational virtuosity) for me to attempt to acquire the painting, or to go and see it in person, or anything else. That is to say, a full theory of the source of practical reasons might not be a full theory of value or of normativity. Korsgaard, though, considers her voluntarist theory a theory about “the source of all value” (\textit{Sources}, 91; also, \textit{Creating the Kingdom of Ends}, 240-241).
Back to Korsgaard’s title: “the normative question.” Why call it “the” normative question? There are other practical standards that admit of what seems to be the same question; intellectual curiosity might lead us to ask an innocent, otherwise unmotivated ‘why?’ of them, too. Korsgaard cannot begin from the assumption that morality is anything more than one practical standard among others. It’s just another voice in the cacophony in our heads, and for all we know, it may be contingent in every sense that the other voices are. Since she does not appeal to the fact of reason, Korsgaard is forced to argue for the claim that morality and its claims are to be set apart from other practical standards—that morality is unique, binding in a special, unconditioned way. That is the task her own beginning assumptions, which are more persistently Cartesian than Kant’s, set for her.

Hence, at this stage in her argument, Korsgaard can call the request for practical justification of morality—of the Moral Law—“the” normative question for no more reason than for the sake of convenience—it is the particular request for practical justification that she happens to be taking up. And maybe for a historical reason—clearly enough, the practical justification of morality is an important question to many of her readers (11). The question ‘Why be moral?’ and the answer ‘Because rationality requires it’ both have a long and storied history, and Korsgaard is adding a chapter.

In any case, we shouldn’t lose sight that Korsgaard’s normative question leads us to a general question about the ultimate source, or sources, of practical reasons. This highly theoretical question about practical reasons is at stake. And she, similar to Kant, doesn’t want us to take a subjectivist theory of practical reasons for granted, as so many attempts to justify morality do. They think we ought to “move beyond” subjectivism, to a theory which includes a “kind of test of normativity” for any desire (91).
The Capacity for Reflection and an Infinite Regress

It is useful to be clear about how the general normative question arises. Our capacity for reflection makes it possible for us to ask for practical justification of morality, or of other purported sources of practical reasons. The capacity to step back from our mental states to ask questions about them distinguishes biologically human beings from biologically non-human animals, and it makes it possible for us to become psychologically very complex. It must be this distinct capacity to call into question our own mental states, and to construct a complex personal perspective, that gives rise to the very common idea that human beings enjoy, within the animal kingdom, an elevated status and are endowed with a special type of dignity.

For all of this elevation and dignifying, the capacity for reflection is also the source of much frustration. (As Descartes could confirm.) Once we ask the questions that reflection makes possible, how in the world are we to go about answering them? The problem is only in part that the questions that reflection could conceivably ask are legion; it’s also that some of the most interesting questions are also so deeply perplexing, even disorienting. The general normative question asks us to call into question even the deepest values and most heartfelt personal motives that we usually simply reason from. Where are we to stand when we ask these questions?

Within Korsgaard’s theoretical framework, the normative question strikes me as even more troubling. If we were to grant the assumptions she brings to the table, the capacity for reflection seems to set for us a chain of questions that cannot be answered.

Think about it this way. If you ask why you ought to treat moral reasons as practical reasons for you, someone might answer that moral virtue pays. But in Korsgaard’s view, that answer does nothing more than make an appeal to another practical standard, prudence, for which
it is possible to request a practical justification, as well. An appeal to another practical standard only provokes another ‘why?’ from restless reflection. Even if you happen to care—deeply and intrinsically—about your long-term interest, Korsgaard is unwilling to treat that as a sufficient, regress-ending answer to the request for practical justification, all things considered.\footnote{For another argument that our capacity for reflection forces us to seek a non-subjective vantage point from which to assess our contingent, subjective conative states, see Nagel (“Ethics”). Nagel, like Korsgaard, considers moral reasoning “inescapable,” though, unlike her, he does not locate the source of the authority of moral reasons in the agent’s will.}

Korsgaard rejects subjectivism; in her view, a desire—even a persistent and strong desire, however inescapably entrenched; or a desire backed by a second-order or highest-order desire—is not itself a source of practical reasons: “the reflective mind cannot settle for . . . desire, not just as such. It needs a reason” (93). Our inclinations simply present for us instances in which we must evaluate and choose: “When desire bids, we can take it or leave it” (93). If we grant Korsgaard’s anti-subjectivism, her Kantian thesis that an appeal to a desire is not a sufficient, regress-ending answer, the natural first impression is that an infinite regress threatens. (Korsgaard’s project is not really driven merely by cases in which an agent feels a bizarre or distressing desire; her question drives much more deeply than that. The anti-subjectivist challenge, in her hands, would only be the tip of the iceberg. She wants to press for questions about every desire.)

Moreover, if we grant that the regress can’t be stopped unless we appeal to something that it is “incoherent, impossible, or unnecessary” to ask a further ‘why?’ about, then the regress begins to build up more speed. Any answer seems to provoke another ‘why?’ from restless reflection. It is possible to ask ‘Why treat that as a source of practical reasons for me?’ of any of a standard list of candidate sources—of an agent’s subjective concerns, of a political sovereign, of an intrinsically normative standard posited by a Realist philosopher, or even of God Himself.
Is there, then, any principled way to end the regress? If we reject each of these candidate sources, there seems to be nothing left to be the ultimate source of practical reasons. Any end to the regress has to come, unsatisfyingly, in the form either of suppressing the questions that reflection asks or answering the question, sooner or later, in an arbitrary way, thereby and thereafter expressing practical deference to something that does not really deserve it. (Worse, actually, in Korsgaard’s theory; you’d be violating an obligation to your reflecting self. Korsgaard thinks we are under an obligation to listen to reflection when it calls out to us—you’d better hope yours isn’t very restless.) For Korsgaard, then, it would seem the only options are to despair of our inability to discover a legitimate and ultimate source of practical reasons or to regard this inability with irony: another example of humanity’s high aspirations setting goals utterly beyond its grasp.\footnote{153} Ask this much of an account of practical reasons and you will be left without.

There is a similarity between the crisis of authority provoked by Korsgaard’s general normative question and Descartes’ quest for certainty in belief. Both have existential starting points: philosophical reflection begins in doubt. Both begin by imagining a human being to be a “bare” thinking thing, a creature with the capacity for reflection, who has lost his confidence in the vantage point from which he previously made his judgments, and so is seeking surer footing. Descartes imagines this thinking thing seeking a firm foundation for theoretical knowledge and belief-formation, Korsgaard for agency and practical reasoning about what to do. Unlike Descartes, Korsgaard does not aspire to provide a “demon-proof” demonstration; she does not demand that our answers to the general normative question be incorrigible. But she does require an analogue. An answer to the normative question has to “speak to” the agent, and to a part of the

\footnote{153 \text{For a discussion of irony, see Nagel ("The Absurd").}}
doubting agent’s very sense of self, his deepest sense of identity. And the answer has to be capable of persuading him no matter how intransigent the spirit into which he has fallen. The only conditions she places upon someone with whom she is speaking about the normative question is that he has to be “reasonable and sincere” and has to “really want to know” (11)—the features of Socrates, and the reflecting self. The right answer has to be such that he can’t shrug off this answer without severe costs, without denying his own rationality and his very self.\(^{154}\) Similar to Kant, the Moral Law is not grounded in either desire or feeling, but in a part of us that is not contingent in the ways our desires and feelings are.

Despite Korsgaard’s high aspirations, she is not driven into despair. She is not a skeptic or a nihilist about practical reasons (94). And she does not take a (mildly) ironical attitude toward the normative question, such as Thomas Nagel might consider.\(^{155}\) She thinks that if we ask, we shall receive, from Kant. She seeks a principled, regress-ending answer to the general normative question, an answer about which it is “impossible, incoherent, or unnecessary” (33) to ask again. And she thinks that Kant provides this answer, or at least he almost manages to. Kant’s theory, she thinks, needs a bit of tinkering, but it is tinkering that can be done, and he has so much of the hard work completed.

The most striking fact about Korsgaard’s view is that the final, regress-ending answer to the normative question, despite her high aspirations, also has welcoming implications for the moral advocate: the right answer to the question whether moral reasons are practical reasons is

---

\(^{154}\) In chapter 2, I compare Korsgaard’s requirement that a theory of practical reasons must speak to even the most intransigent agent with the “internalism” requirement, and Hubin’s talk of “shruggability.”

\(^{155}\) See “The Absurd.” Korsgaard can’t accept Nagel’s strategy to defend a type of realism. Nagel thinks that once the various objections to realism are met, there isn’t any special reason to doubt its truth. No positive arguments need to be brought to defend it. Korsgaard wants her argument to speak even to the person who “doesn’t see it just now,” who lacks Nagel’s “polite” sense of confidence.
‘yes’, and universally so. For each human being—no matter what he displays practical deference to—if he were to ask the question, the answer would be ‘yes’. Subsequently, the moral philosopher is able to promote the moral life with a good conscience, simply by telling the truth. We can be “transparent” about “what morality is and why we are susceptible to its influences” (17). (Thank goodness for Kant. Given his absolute prohibition against prevarication, morality would forbid even a noble one.) According to Kant and Korsgaard, the universal authority thesis comes out true on the right understanding of the ultimate source of practical reasons, and is firmly grounded in “the unconditioned” (94). Immoral inclinations are always properly disenfranchised; the will, as practical reason, has the right to extrude any immoral desires from the context of practical deliberation. More, the will not only has the right to, it ought to. Whether the agent recognizes it or not, an immoral desire does not deserve a voice.

Kant’s theory, to say the least, is strikingly different than my refined subjectivist theory. Recall in the first chapter I mentioned that many Kantians think that intuitions such as the intuition that the will has both the capacity and the practical authority to disenfranchise some desires should lead a person to consider Kant’s more robust theory of the will. It is difficult to see how the will, as construed by a subjectivist such as Frankfurt, is able to rule out desires. How do desires rule out desires? How do desires—call them “ruling passions, if you like”—issue commands? Barbara Herman, for instance, thinks that anyone with Frankfurt’s general intuitions ought to consider “the compelling attraction of Kant’s theory” of the will. In the first half of the dissertation, I attempted to refine Frankfurt’s account of the will, in a theoretically modest way to help Frankfurt overcome the challenge. In what follows, I will take up Herman’s invitation to consider Kant’s theory of the will. But I don’t find it quite so attractive.
In Korsgaard’s case, the general argument that Kant makes takes a detour through an appeal to an agent’s normative self-conceptions, and she regards a normative self-conception as a practical standard that includes norms for treating desires as reason-giving, including what I would describe as norms of rational impotence. Korsgaard thinks that it is necessary, a rule of rational willing, that a reflective agent self-legislate at least one practical identity, though most people have “a jumble” of them. Her proposal resembles mine in several important details. Her ‘practical identity’ is very similar to my ‘personal ideal’. But she thinks a personal ideal derives its practical authority to guide action and deliberation from being self-legislated, independent from desire. Whatever ways Korsgaard differs from Kant, she continues to treat the will as strangely alienated from the deepest concerns of an agent’s heart.

What is Kant’s answer? It has to do with the Categorical Imperative. Kant tells us that the Categorical Imperative is a rule of rational willing, a rule of personal autonomy, and the Moral Law. One and the same principle, three purposes. Kant’s trinity. To obey the Categorical Imperative is to be practically rational, personally autonomous, and morally dutiful, all at once. (At least, of course, if you employ the principle lucidly, that is, with the full light of reflection turned on.)\(^\text{156}\) And to violate it is to be, to one degree or another, practically irrational, personally unfree, and immoral, all at once. In this way, Kant’s theory bears—at least at a very general level—a striking elegance. Kant attempts to bring together a set of three highly complicated notions—rationality, freedom, and morality—into a surprising and exquisitely unified general theory of rational and moral agency.\(^\text{157}\)

\(^{156}\) Kant is confident that any person is able to discern the Moral Law and what it expects of us (391).

\(^{157}\) Kant thinks that rationality, freedom, and moral duty are all exemplified, in a deeply unified way, by the fully rational agent. What a fully rational agent would, of necessity, do is what you and I ought to do.
Soon enough, in the next chapter, we will see that the striking elegance of Kant’s theory at one level masks a mind-boggling complexity at another. But let’s ignore the underlying complexity for a few moments more, to focus our gaze upon the exquisite surface. It is useful to begin with Kant’s rational anthropology in which he “defines” what a human being is and makes a connection between the first two elements of the trinity, between rational willing and personal autonomy. Following the rules of rational willing, Kant thinks, is personally liberating in two especially deep and important senses.

Practical Rationality and Personal Autonomy

Kant defines a human being as a rational creature. He thinks that, in essence, a human being has the faculty of reason, which is comprised of both theoretical and practical reason. Otherwise put, a human being has an intellect and a will, which means she is capable of thinking about whether her beliefs are objectively true and her willings are of genuine objective worth.

Kant believes that there are rules of rationality, of both theoretical and practical reason. This has important implications for Kant’s “rational anthropology” and his theory of rational agency. Since the capacity for rationality is a part of our human essence, Kant thinks that the rules of rationality are a part of the very fabric of our human nature. Therefore, according to Kant, our human nature determines, to a degree, how we ought (not) to behave; it is the source of rules that constrain how it is proper for us to think and to act. Human nature itself, through reason and its constitutive rules, gives us a measure of practical guidance.\(^\text{158}\)

\(^\text{158}\) This means that Kant’s theory shares something in common with teleological conceptions of practical reason, such as Aristotelian realism. But with an important difference. Both outlooks posit a human nature that is normative for us. But the substantive content of human nature is differ in the two accounts. Broadly speaking, Aristotelian conceptions of our telos are substantive, and the substantive rules for a good human
Also, in Kant’s view, the rules of rationality make up an important aspect of an agent’s self. The rules of both theoretical and practical reason are, for us as rational beings, self-constituting rules. This takes us again to Descartes’ influence. Kant’s conception of the self is not merely of a so-called Cartesian ego: a thin, wispy, reed-like capacity for reflection. The rules of rationality “thicken” the self; they provide, Kant thinks, a lasting foundation from which to make decisions about what to believe and about what to do. As empirical data and subjective impulses stream into our conscious life—into the citadel of the self, we might say—these rules are able to help us to decide, properly, what to believe and what to do. It is useful to speak of Kant as thinking that each human being has a practical point of view, and the rules of rationality are a part of it, and as they operate on the raw material that our experience as embodied creatures with needs and desires presents us, they begin to give further shape to our developing practical point of view.

Kant’s system is also committed to an aspect of the natural law tradition: there are rules for behavior that are “built into the fabric of things”—into our nature. Kant’s system is also committed to an aspect of the natural law tradition: there are rules for behavior that are “built into the fabric of things”—into our nature.

Kant consistently incorporates into his own theory elements of the philosophical traditions he is otherwise rejecting. His argument seems to attempt to treat many traditions as each having a kernel of truth that he intends to steal way. In this way, Kant’s theory includes elements of realism, the divine command theory, and the natural law tradition. His theory is realist in the sense that the rules of practical reason he posits have the type of independent existence that realist-type rules are generally considered to have. They simply exist, as a part of “the very fabric of things”; there are rules of theoretical and practical thinking. But his theory is a form of “procedural,” not “substantive” realism, which is constructivist in nature, because these realist rules do not tell you how (not) to project yourself into the course of events in the mind-independent world, they are rules for how to properly construct a set of rules for the governance of life. Practical reasons derive from the principles of choice that pass the test of practical reason. These practical reasons are “constructed” out of the activity of applying the procedural rules of practical reason to the maxims associated with an agent’s inclinations to act (112). And this is all a form of voluntarism, too, because you self-legislate the rules that thereby become practical laws for you. They are the consequences of actions.
These ideas have a profound influence upon Kant’s conception of freedom. Kant’s theory of personal autonomy emphasizes self-governance. A self-governing human being is governed, through his intellect and through his will, in accordance with rational, self-constituting principles. To be guided by the rules of theoretical and practical reason, since they are aspects of the self, is to be autonomous—or self-governing. We behave freely by following our nature, which is rational. Since the will is practical reason, a fundamental aspect of the very self, following it is autonomous. On the other hand, a person lacks personal autonomy if she is governed by—or pushed around by—“heteronomous” masters. Kant considers behavior that is ultimately caused by anything “external” or “alien” to the agent’s will to be heteronomous. And he considers several distinct sources of influence heteronomous: the commands of other people, or even of God; the various practical standards we commonly encounter in the flow of life, such as prudence or etiquette or a social code of conduct; and the person’s own subjective desires. You are not acting autonomously if what prompts you to act, ultimately, is mere practical deference to (a) another person who has issued you a command or (b) a practical standard. Also, and importantly for my arguments, in Kant’s view, the contingent, subjective conative states of an agent are “alien” to the agent’s self and to her will. (Call this the “Heteronomy Thesis.”) You are not autonomous if what leads you all the way to action is, ultimately, an impulse such as fear of punishment or a desire for happiness, or simply the idea “Because I want to.”

---

of a will, yours. The rules of practical reason are constitutive of the defining aspect of your nature and your self, and so, similar to the natural law tradition, at least in its Aristotelian manifestation, the relevant rules for how to behave are built into your nature.
According to Kant, an intentional action is done from a principle and not from a desire.\textsuperscript{161}

Kant generally speaks of an agent’s impulses to action, anthropomorphically, as “bidding” the will to act in the way that they suggest. An impulse “presents itself to” or “presses itself upon” the will (\textit{Critique of Practical Reason}, 74)—and it does so “through a maxim” (a subjective principle of action). To be free, a human being must subject these maxims to tests, and they are to be followed—acted upon—only if they are endorsed by the rules of practical reason. Kant means to suggest, also, that the human self is therefore not merely a passive observer, but also an active thinker and an active agent, a decision-maker about what to think and about how to act. The self, according to Kant, is a legislator, and the rules of rationality provide him with a vantage point from which to make decisions.\textsuperscript{162}

Kant’s definition of ‘human being’, we should be clear, is technical. It is not intended to track common usage. Kant means for the definition to pick out all and only those beings that make up the moral community, and all and only those to whom the Categorical Imperative, and so the Moral Law, is supposed to apply. This community includes any being with the capacity for reason, regardless of whether the being is genetically human, and excludes any being that lacks this capacity, even if it is genetically human. Clearly, Kant supposes that any normal, adult human being is included in the class of human beings in the technical sense. But his argument is supposed to reveal that the Moral Law is valid for any being that has the capacity for reason no matter what else is true of him. We should make sure that Kant sticks to his own definition.

\textsuperscript{161} Kant does not have an artificially constrained set of actions he considers voluntary. I think it is fair to say that Kant takes most of the actions that we commonly judge to be voluntary to be so.

\textsuperscript{162} As sense data provide the passively received material that the mind then actively organizes into a consistent theoretical outlook, so desires provide the passively received input that the will then actively organizes into a consistent practical outlook.
When a philosopher stipulates a new and technical meaning to a common term, the danger arises that, at some later stage in his argument, assumptions connected to the old and non-technical term slip in. Kant, I think, succumbs to this danger.

Two Aspects of the Will

I would like to adopt a few locutions, to help keep my discussion of the will clear. The rules of practical reason (that is, of rational willing) are not the only rules that make up a particular human being’s will. A human being is able to act freely by guiding his own conduct in accordance with self-legislated principles of action—“maxims”—that reflect external standards of conduct or her own contingent, subjective concerns. But the agent will be free in doing so only if these self-legislated principles are consistent with the rules of rational willing. It is the specially ordained work of the will—its natural function—to decide whether the subjective principles “associated with” the agent’s impulses to act are fit to be acted upon, and this is a matter of their being consistent with the rules of practical reason. If a maxim does pass this test, it can be incorporated into the agent’s will.

So, I think it is helpful to see Kant as thinking that the will—the self-constituting practical point of view—of any particular human being consists of two sets of principles: (a) the rules of rational willing (the rules of practical reason itself), which are valid for the agent, as such, in virtue of his rationality, and (b) the self-legislated, subjective principles (maxims) that reflect the particular agent’s subjective, desired ends. To provide an example of each type of principle, the principle that I ought to will the indispensable means to the ends I have set myself to pursue is valid for me as a rational creature, and it is valid for any rational creature, as such. But my
personal policy to play soccer whenever it doesn’t put out my wife, Kant would say, reflects one of my particular material ends (a desire or intention to play soccer), and is valid only for me, given my desired end, and not for any rational creature, as such.

For the sake of clarity, it will be useful to have titles for each aspect of the will. I am going to call the various rules of rational willing aspects of the agent’s “rational will,” and an agent’s various self-legislated, subjective principles jointly make up his “personal will.” A fully rational agent will have a will that consists of the rules of rational willing, along with self-legislated principles of choice that are endorsed by the rules of rational willing. The elements of an agent’s rational will are universally valid and rationally inescapable; the rules of her personal will are not. This general view implies that human beings are both importantly similar and importantly different from each other: we are all subject to the same rules of rational willing, but might have very different self-legislated principles of choice that reflect our different personal tastes and goals. There is, in a good human community, both diversity and profound grounds for solidarity.

One more remark. Kant considers the rules that make up an agent’s rational will as, we might say, “rationally inescapable.” What does this mean? In violating a rationally inescapable rule you violate your very rationality and so your very humanity. Some maxims, Kant thinks, are inconsistent with the rules of rational willing. Hence, your rationality and humanity set limits upon how you are permitted to construct your personal will. These rules are given, not chosen,

---

163 To foreshadow, in chapter 5 I will raise the question whether Kant thinks there are some rules of rational willing grounded not in our rationality, but instead in other aspects of our humanity. That is, I will ask whether our humanity is merely a matter of our rationality, or whether there are other aspects that make us human. Oftentimes Kant speaks as though our rationality is fully constitutive of our humanity. But to protect him against an objection, I will suggest that that our humanity might consist in more. See the section concerning the fourth objection to Kant.
and so provide a kind of stability to our practical point of view. Whatever else might change about us—our personal priorities, needs, hobbies, and so on—these remain the same.

Two Rules of Rational Willing and Two Types of Freedom

What’s at stake between Frankfurt’s subjectivism and Kant’s theory is who does a better job of capturing the intuition, mentioned in chapter 2, that we do confront ‘musts’ in practical reasoning, ‘musts’ other than the ‘must’ issued by the instrumental principle. Frankfurt, of course, intends to capture this intuition without including in the rational aspect of the will anything more than the instrumental principle and other considerations of consistency. Our ruling passions are supposed to do the work of creating these various ‘musts’. My proposal attempts to capture one of them: the sense that some basic desires cannot properly be treated as reason-giving. Kant appeals, instead, to more or “extra” rules of rational willing to be incorporated into the rational aspect of the will.

Kant proposes (at least) two basic rules of practical reason: the Hypothetical Imperative, which is a basic rule of instrumental rationality, and the Categorical Imperative. To govern conduct by the Hypothetical Imperative and the Categorical Imperative is to be both practically rational and to have a free will. The Categorical Imperative is a rule of rational willing; and since we are in essence rational creatures, it is also for us a principle of personal autonomy.

The fully rational agent displays two types of freedom. First, personal autonomy. A fully rational human being “makes laws unto himself.” He constructs his own practical point of view—through acts of volition, guided by practical reason. In this way, he sets his own path for life. He owns the personal aspect of his practical point of view in a special sense. With respect to
it, he is a “self-made man,” not a slave to convention or to his own natural endowment. (However niggardly.) With respect to his rational will, it is a given and not constructed by him, but this doesn’t compromise his freedom because the principle he is following—the Categorical Imperative is also the “Principle of Autonomy.” In following it, he is affirming his own rational human nature. (And he is no more properly described as a slave to this principle than a person who reasons validly is properly described as a slave to logic.)

Second, a fully rational agent is not self-divided, an obstacle to himself. He is free of division within the will. The world might present him with difficult obstacles in his attempt to pursue his desired ends, but he does not have to overcome obstacles within his will, within his own practical point of view. For one paradigm of unfreedom is the person who adopts inconsistent principles of choice and subsequently violates one in following the other. This inconsistency pulls him in two directions, so that no matter what he does, he frustrates one of his own intentions; or he ends up stuck like Buridan’s ass, doing nothing. Since the rules of rational willing enjoin an agent to construct a practical point of view that includes only principles of action that are consistent with each other, he doesn’t suffer from this type of unfreedom. Kant thinks practical rationality requires that an agent will only maxims that are consistent with each other, and consistent with the rules of practical reason itself.

The Moral Law

So far, I have connected, in broad strokes, the first two members of the trinity. What about the third? How is it that the Categorical Imperative is also the Moral Law? This proves to
be the tricky question, and it will exercise me for some time in the following chapter. But a gloss is sufficient for the moment.

Kant stresses consistency. As mentioned, Kant thinks that the contingent, subjective concerns of an agent are external to his will, so an autonomous being cannot be governed by them. Kant thinks that the Moral Law is valid for all rational beings, and so its authority cannot be grounded in the contingent subjective states of an agent. (If it were, the Moral Law would not be rationally binding upon any rational creature who happens to lack the relevant, authority-grounding desire.) Instead, a free will follows principles. Kant infers that the supreme principle of morality must be a formal principle, grounded only in the rules of rational agency and the conditions that are necessary to make rational agency possible. It has to be a rule that binds a rational agent, as such. Kant thinks formal consistency is a requirement of rationality, as such. So, the Categorical Imperative is a formal principle that requires consistency in willing. The subjective principles that we self-legislate must meet the standards of formal consistency.

Again and again Kant insists that the Categorical Imperative is, and must be, completely formal. Kant also regards logic as the only thoroughly formal science (387-388). Also, when Kant discusses applications of the Categorical Imperative, he calls it the “principle of contradiction” and the “merely formal condition” of “thoroughgoing consistency” (422, 424). Violations of the Categorical Imperative are “contradictions” in the will (422-424).

Kant therefore indicates that he seeks a law for practical willing that does the work that the logical law of non-contradiction does for belief. The logical law of non-contradiction rules out any proposition of the form ‘P and not-P’, and it suggests a principle for belief-formation: ‘Do not believe both P and not-P, or things that imply both P and not-P’. This principle does not
tell you which particular propositions to believe; it tells you not to have self-contradictory beliefs or an inconsistent belief set. Kant thinks that the logical law of non-contradiction is necessary for the very existence of theoretical reason (Critique of Practical Reason, 120; Lectures on Ethics, 51, 55-57). It is a part of the “logic” of theoretical rationality. Likewise, it would seem, Kant thinks the Categorical Imperative is a formal principle that enjoins an agent to have neither self-contradictory volitions nor an inconsistent volitional set: ‘Do not will both W and not-W, or things that imply both W and not-W’. This principle, Kant thinks, is necessary for the very existence of practical reason, and so is a part of the “logic” of practical rationality. Since following the logical principle of non-contradiction makes rational thinking possible, following the Categorical Imperative similarly—according to Kant—makes rational action possible.

What is striking is that Kant thinks that the Categorical Imperative, though a formal principle, has substantive moral implications, and that these implications are rationally inescapable. The Categorical Imperative is able to rationally compel, Kant thinks, respect for the dignity of every human being and for social equality. Any agent that fails to respect the dignity and equality of all human beings (i.e. rational beings) suffers from an irrational form of volitional inconsistency and is violating the tenets of his own nature as a rational being.

How does it all fit together? Kant’s argument happens in two general steps. First, Kant argues from reflection upon rationality and the conditions of rational agency to a formal principle of rational willing, which Kant will call the first formulation of the Categorical Imperative, and then from this formal principle to two further principles, each of which are, he tells us, formulations of the Categorical Imperative and substantive moral principles.
Korsgaard’s General View

Korsgaard agrees with much of what I have attributed to Kant. Korsgaard advocates Kant’s rational anthropology. She defines a human being as, in essence, a creature with the capacity for reflection and for theoretical and practical reason, and she treats the will and the rules of rational willing as (partly) constitutive of the self (along with the intellect and the rules of rational believing). She agrees with Kant’s conception of personal autonomy and freedom, which includes the Heteronomy Thesis and the claim that the will, if it is to be free, must govern through principles. A desire, as such, has to pass a “kind of test of normativity”—a test of practical reason itself—to be able to play a proper role in the generation of practical reasons. Also, Korsgaard regards the demands made by a political sovereign or by a preacher or by God himself as alien to the agent’s will; they are, if treated as masters, heteronomous. To be free—in the sense of self-governing—an agent must act according to self-legislated principles that meet the principles of volitional consistency. Consequently, she also accepts, implicitly, the two aspect theory of the will, according to which the will is made up of the rules of rational willing and properly self-legislated principles of choice.

Korsgaard accepts a Kantian normative ethical theory, that is, a Kantian conception of content of the Moral Law, with its emphasis on respect for persons and on fairness and social equality. Her normative ethical theory is also deontological, and she is searching for the ultimate source of moral obligation. Morality issues obligations because we are “reluctant, recalcitrant, resistant matter” (4): we often do not accede to morality’s demands without a struggle. Also, she advocates one of Kant’s several conclusions. She argues for the universal authority thesis: moral reasons are pro tanto practical reasons for any practicing agent, as such.
The structure of her argument is also Kantian: beginning in with a Cartesian question, she appeals in ways similar to Kant to rules for volitional consistency and the constraints they are supposed to place upon rational and free behavior. These constraints will force us, if we are to fulfill our rational nature, to the path of a life of moral duty. According to Korsgaard, the source of moral obligation is, at the very same time, the basic principles of rational willing and of personal autonomy—Korsgaard agrees with Kant that the rules of rational willing are also rules of personal autonomy, and the Moral Law is derivable from them.

Disagreements with Kant

But Korsgaard also has important disagreements with Kant. Korsgaard’s view, though clearly Kantian in spirit and in many important details, self-consciously departs from Kant in at least one important respect, which Korsgaard regards as necessary. In Kant’s view, each impulse to act has to pass a test of reflection, the Categorical Imperative, which—Kant believes—simply is the Moral Law. Korsgaard regards the Categorical Imperative and the Moral Law as distinct and so separates the members of Kant’s trinity. According to Korsgaard, the Categorical Imperative is—as Kant believes—both a rule of rational willing and a rule of personal autonomy, but it is not the Moral Law. And so, Korsgaard thinks that it is necessary, if she is to soundly defend the universal authority thesis, to argue from the Categorical Imperative to the Moral Law.

Why, according to Korsgaard, doesn’t Kant’s argument work? The primary objection Korsgaard expresses is that Kant’s Categorical Imperative, being a formal rule, does not have sufficient content to rule out every immoral principle an agent might self-legislate. It is in this

---

\(^{164}\) Sullivan (Immanuel Kant’s Moral Theory, chapter 11) and O’Neill (“Consistency in Action”) disagree, advocating that a maxim is a highly general “life plan” that governs day to day decisions, whereas the Categorical Imperative governs which life plans are obligatory, permissible, and forbidden.
sense, as it is often called, an “empty formalism.” This would mean that Kant’s theory is in need, in the very least, of supplement: the formality of the Categorical Imperative must be supplemented by something not thoroughly formal if we are to soundly argue from the rules of rational willing and personal autonomy to the Moral Law.

In short, Korsgaard thinks there is a gap in Kant’s argument, and she takes it upon herself to fill this gap. Her argument appeals to the concept of a “practical identity”: an action-guiding self-conception. Korsgaard argues, first, that any creature with the capacity for reflection must guide her conduct by a practical identity and, second, that any creature that does guide her conduct by a practical identity is rationally committed—by rules of volitional consistency—to the Moral Law. At least upon one reading, Korsgaard substitutes talk of how a practical identity robustly shapes an agent’s practical deliberation for Kant’s talk of maxims and their formal evaluation. We should anticipate, then, that a practical identity is not a purely formal standard; whatever a practical identity is, it includes substantive rules, among other things, for conduct. Korsgaard also considers a practical identity to be a test of reflection for inclinations.

For Kant, it is the Categorical Imperative itself that evaluates the claims made by a desire. It is a norm of rational impotence. For Korsgaard, the norms of rational impotence are contained in the action-guiding self-conceptions that agent’s have properly self-legislated.

Beginning Assumptions

There are several other features of Korsgaard’s argument that are importantly distinct from Kant’s. Kant’s argument begins from a sense of confidence in the Moral Law. He thinks we all already recognize, at some level of consciousness, the validity of the Moral Law; it is a
part of “natural sound understanding” (407), and “reason presents” it as “so deserving of respect” (405). Kant is the St. Paul of moral philosophers (and the *Groundwork* is his Letter to the Romans). He thinks that if we behave immorally we are “without excuse.” For a particular person, the sublime character of the Moral Law might be half-hidden, behind veils of self-deceit and personal anxiety, but Kant is convinced that such a sense is implanted in the human mind and strongly enough to make us culpable for misbehavior. This moral sense is the analogue to the *sensus divinitatis* posited by St. Paul. Along with the starry heavens above, the Moral Law “within” provokes profound respect, even awe. Kant calls this intuition “the fact of reason.” (And he asks what we and the Moral Law must be like for this to be true. Confident that the Moral Law does deserve respect, Kant shapes, it seems, his anthropology in accordance.)

Korsgaard doesn’t appeal to the fact of reason. She doesn’t claim that the Moral Law receives an endorsement from reason’s intuition. She doesn’t claim that we know in our mind of minds, by a dispatch from intuition, that the Moral Law is valid; some people might really need convincing. Human beings don’t necessarily have an implicit trust in conscience, or if we do, we can’t rightly appeal to it. She’s more “Descartes in the 1st Meditation” at this point; reflection can call any sense of certainty or confidence into question easily enough. A person who feels the force of conscience might ask whether it has a right to tell her what to do, and with respect to its authority, she “might not see it right now.” The question reflection is asking is really troubling her. (This is what leads Korsgaard to reject the realist’s route to secure the universal authority thesis; it simply posits that moral reasons are practical reasons, and so has nothing to say to the person who has lost confidence in the validity of moral reasons.) It follows, then, that Korsgaard must resolve to rely exclusively upon the soundness of her argument for the universal authority thesis.
So, Kant and Korsgaard both think that it is a good idea for people to learn the argument for the universal authority thesis. Kant thinks, though, that people need more to be reminded than taught that the Moral Law deserves unconditional respect. Korsgaard, it would seem, has to say that there are people to whom this really needs to be taught. The ‘all’ in the claim that “all of us confront the normative question” refers, coming from the mouth of Korsgaard, to those of us in the modern west, influenced by the Enlightenment. All that Korsgaard asks of someone who is following out the normative question is that she be “sincere and reasonable” and that she “does really want to know” (16). That is, she only has to be asking the question from the perspective of innocent, impartial, curious reflection.

But setting this aside, Korsgaard might well have an advantage here. For Korsgaard is not committed to Kant’s Pauline claim. Since Korsgaard departs from Kant in that she does not claim that we implicitly recognize the validity of the Moral Law, Korsgaard could be construed as a Kantian chastened by a greater awareness of cultural diversity in ethical opinions. Achilles, then, might not have confronted a counterpoise between a sense of fairness to all members of humankind and his heart’s desire. He wouldn’t have framed things that way. And Korsgaard doesn’t have to say that, at some level of consciousness, he did. Korsgaard doesn’t have to answer the objection raised earlier against Kant that our experience of the Moral Law is culturally determined and not an essential aspect of human experience.

All this said, Korsgaard shares Kant’s ambition. She intends to argue for the Universal Authority Thesis, and it is supposed to apply to any being with the capacity for reflection, even Achilles. She has to say that there is some route or other by which we could reveal that the Moral Law is a law even for Achilles, or any person. She is committed to the view that any human being, having the capacity for reflection, is subject to the Moral Law. The resources for revealing
the authority of the Moral Law are contained within the capacity of reflection and the structure of
the human mind, no matter cultural diversity. Subsequently, Korsgaard does think that if we—if
any being with the capacity for reflection—really thought about it, if we really follow out the
questions that reflection naturally raises and “push reflection as far as it will go,” we would come
to see this. Korsgaard shares Kant’s view that we can catch someone who has misbehaved
morally in a contradiction, even if it isn’t true that he is intuitively aware of this. For Korsgaard,
the validity of the Moral Law might be hidden in the recesses of difficult chains of reasoning, and
not immediately clear to common sense intuition. But it’s there, for a sincere, clear-thinking,
curious reflective agent to find.

Put historically, in Korsgaard’s view, the Enlightenment worldview, in at least some of
its attitudes, and so Kant, in at least some of his, is legitimately privileged. Socrates sowed the
seeds of the Enlightenment project. He asked us to reflect upon our ideas of what the world is
like and our ideas about how to live, to see whether they stood up to scrutiny. As Korsgaard sees
it, the Enlightenment project took up this question full force, and made important progress. It
reveals what was already latent in the questions Achilles did ask (whatever precisely they were).
In her view, it also hit upon the right values: liberal, broad-minded concern for respect and
fairness.

Korsgaard does, though, think she gets help from a strong and trustworthy common sense
intuition. Whereas Kant posits the fact of reason, Korsgaard posits “the fact of value.” This is
the sense that there are legitimate practical standards: “It is the most striking fact of human life
that we have values” (1). We all feel it in our psychic bones that some things really do matter, are
“normative” for us. This is what we all confront; Achilles, too. He felt this profoundly; he had to
choose to give up one thing that mattered to him, his life, to protect another, his honor.
Reflection is able to call any idea about how to live or how to be into question. But Korsgaard thinks that we aren’t likely to doubt that there really are legitimate answers. Korsgaard calls the fact of value, our sense that there really are legitimate values, a “mystery” (2). And she says that “philosophers have been trying to solve it ever since” Plato made Socrates “see the problem.” In other words, Korsgaard thinks the fact of value is a feature of life that we share even across an intellectual divide as great as that created by the intellectual revolution that brought us from the ancient and medieval world into the modern scientific worldview. Even though the world was “turned upside down and inside out” in this revolution, we share with the ancients a sense that there really is a problem; we have to adopt a view of the world, a “metaphysic,” that makes sense of our commitment to the existence of “value.” Korsgaard pays homage to Plato by indicating that his theory at least gets the phenomenology of value right. We experience what some have called “the magnetism of the good”:165

where do we get these ideas that outstrip the world we experience and seem to call it into question, to render judgment on it, to say that it does not measure up, that it is not what it ought to be? . . . it is puzzling too that these ideas of a world different from ours call out to us, telling us that things should be like them rather than the way they are, and that we should make them so” (1).166

Korsgaard trusts that if we were forced to choose between the view that there aren’t any values whatsoever, and subsequently, a life without any value commitments whatsoever—what she calls ‘practical nihilism’—on the one hand, and moral identity, on the other, we would—in part because of the “fact of value”—choose moral identity. She takes it upon herself, then, to reveal that these two options are the only real options. That’s the dilemma she needs to force us into, by

165 For instance, see Hare (God’s Call, 13-14).

166 Korsgaard remarks, further, “Clearly, we do not get them from experience, at least not by any simple route” (1). (A big “at least”!)
argument. (Since we cannot, she thinks, return to the ancient or medieval ways of unlocking this mystery, we’ll have to come by an answer consistent with the modern scientific world view, an answer which is a voluntaristic, constructivist, procedural realism.)

Korsgaard aims for a weaker set of conclusions about the authority of moral reasons than Kant. Let me clarify the conclusion Korsgaard is arguing for, what it is and what it’s not. The universal authority thesis is the claim that moral reasons are \((pro \ tanto)\) practical reasons for every rational creature. But, you might wonder, what is going to count as a “moral” reason? The Moral Law, as construed by Kant, is very complex. Among other things, it forbids, absolutely, lying and suicide, and it demands respect for the dignity of other human beings, where respect amounts to giving normative weight in your practical deliberation to the needs and proper concerns of others. It is only this last aspect of Kant’s broader conception of the Moral Law that Korsgaard is particularly interested in defending. Even if her argument did not vindicate everything Kant includes under the Moral Law—an absolute rule against lying, say—Korsgaard tells us that she would see her argument as vindicated if it soundly argued for the conclusion that you and I are under a rational requirement to exemplify respect for other human beings.

Also, Kant thinks that moral reasons always trump other types of practical reason, that is, that moral reasons are “supreme.”\(^{167}\) Korsgaard is not aiming to defend this claim. She is attempting to argue for the weaker claim that moral reasons always carry some weight in practical consideration, not that they are always the weightiest consideration. Or, to use another metaphor, moral reasons always deserve a voice, but not always the last word. In other words, the “authority” that the Universal Authority Thesis attributes to the moral reasons is not final or

\(^{167}\) See Brink (“Kantian Rationalism”).
overriding authority. It is enough, she thinks, to get to the conclusion that moral reasons are universal, pro tanto practical reasons, even if not to the conclusion that they are universal and universally final. I think it is fair to say, though, that Korsgaard aims at a slightly stronger conclusion than this. She thinks moral reasons deserve, to put it vaguely, a central or important place in practical deliberation. I suspect Korsgaard’s view is that (weak) moral considerations are sometimes outweighed by strong non-moral but never—or rarely?—by immoral practical reasons. This general description, I hope, is sufficient for the time being; I’ll say more about it in the section about Korsgaard.\footnote{Korsgaard also does not mean to defend the “unity” of the Moral Law, the idea that moral obligations never conflict with each other in a deep way. Kant seems to have thought that the voice of morality, so to speak, is always singular and wholehearted. Korsgaard is willing to entertain the idea that morality might issue inconsistent obligations.}

Also, Korsgaard’s taking off point is an existential loss of confidence, not a sense of the sublime authority of the Moral Law. Korsgaard begins from the sense of practical disorientation a person afflicted by restless reflection might feel when she does not have a sense of her own identity. This leads Korsgaard to an emphasis upon staving off practical disorientation and upon questions of existential meaning.\footnote{See Herman’s discussion of Korsgaard ("Bootstrapping").} In fact, Korsgaard’s argument turns on the thought that the adoption of principles of choice, in particular the adoption of a practical identity, staves off the death or diminution of “the self.” There is a sense in which, in her view, practical reason itself does not provide sufficient continuity to an agent’s life. It needs help, in Korsgaard’s view, from our self-conceptions. (In this way, her view has similar themes as Harry Frankfurt’s, which is explicitly and self-consciously non-Kantian in several respects.) And Korsgaard admits that this emphasis upon self-conceptions introduces “a bit of relativism” into her system of thought. Our
self-conceptions are contingent and often culturally determined, but she insists that they “really”
do obligate us; and in her view, they do so even if the Moral Law condemns them. Korsgaard’s
existentialist language and points of emphasis aren’t, in the very least, characteristically
Kantian. And we’ll have to examine how she fits this “relativism” together with the universal
authority thesis.

The Goals

In the following two chapters, we’ll examine Kant’s and Korsgaard’s proposals for how
to account for the will’s capacity and practical authority to disenfranchise a desire and the claims
it makes. Kant, I will argue, treats as rules of rational willing several principles that do not
deserve that status; and though Korsgaard attempts to escape several common objections to
Kant’s theory and argument, her theory fails to give a coherent theory of normativity. Her theory
leads to an incoherent “chain of command”; she separates the reflecting self and the legislating
self and gives ultimate practical authority both of them—or so I will argue.

Before I describe Kant’s argument for the universal authority thesis, it is worth remarking
for the sake of clarity that Kant’s treatment of matters ethical is systematic. And so, his theory
has provoked a wide variety of objections. For instance, various commentators claim that Kant’s
normative ethical theory, among other things, puts too much emphasis on duty and ignores the
special place of friendship, love, and affection in the moral life. I am not going to be discussing
these objections—or many other objections to Kant’s general theory—at least not directly. I am
not concerned here with the content of Kant’s normative ethical theory, but instead with the

---

170 For similar remarks, see Nagel’s commentary (Sources, 200-209).
several Kantian claims about the ultimate source and character of practical reasons, the claims Korsgaard hopes to vindicate. Most broadly, I will challenge the claim that practical reasons derive from principles for action and not from desires, and the (voluntarist) claim that the principles for action that make up the human will are the ultimate sources of practical reasons. Also, I will challenge the claims that the Categorical Imperative is a rule of rational willing, and that the Moral Law (properly construed) can be derived, at least with Korsgaard’s supplement, from the appropriate application of the Categorical Imperative. These are the claims that Korsgaard is seeking to defend by supplementing Kant’s general theory.
CHAPTER 5
KANT AND THE CATEGORICAL REJECTION OF DESIRE

It goes without saying that Kant’s influence on theorizing about morality and practical reason has been profound. H.J. Paton remarks that Kant’s *Groundwork of the Metaphysics of Morals* is “one of the small books which are truly great: it has exercised on human thought an influence almost ludicrously disproportionate to its size” (“Preface”). Much of what I say about Kant is taken from the *Groundwork*, for his primary purpose in writing it was to “seek out and establish” the supreme principle of morality, the Categorical Imperative, “to formulate that principle and to reveal that we are bound by it” (392). So, the *Groundwork* is the natural place to find much of Kant’s argument for the universal authority thesis, the claim that the Moral Law is a rule of rational willing, an “objective law,” valid for all practicing agents, as such.

Textbook treatments of Kant often fail to capture the power and subtlety of his image of human life in a moral community. I don’t want to produce a flat and uninspiring distillation of Kant, but my focus is not on capturing the beauty of Kant’s moral vision. I will evaluate Kant’s highly analytic argument for the will’s authority, “as practical reason,” to disenfranchise desires. I will be doing the work of what Hume calls the “anatomist,” dissecting arguments, and not the “painter,” making the moral life seem maximally appealing.

---

171 For a discussion, see Sullivan (Immanuel Kant’s Moral Theory, 149).
I don’t think Kant’s theory of practical reason is as compelling as his moral vision. I will argue that Kant, to get to the universal authority thesis from reflection upon the nature and conditions of rational agency, includes many implausible assumptions about the rules of rational willing. One of them: the Categorical Imperative is not tenable as a universally valid norm of rational impotence, a rational test for (basic) desires.

The Categorical Imperative and the Moral Law

According to Kant, the Categorical Imperative, and so the Moral Law, is a single law that can be formulated in several ways. Kant says three (436). These three formulations are representations of the same law, Kant thinks, in the sense that they each lead to the same practical judgments; but they are different in the sense that they bring out different aspects of the Moral Law (436). The second and third formulations are especially helpful for moral education; they make the highly abstract first formulation, which is to be seen as the practical principle of non-contradiction, more appealing (404), and its meaning more transparent and compelling to his readers, “closer to intuition” and “nearer to feeling,” as Kant puts it (436-437).

Some commentators have claimed that there are more: Pogge (“The Categorical Imperative”), and Sullivan counts twenty variations of the first formula (Immanuel Kant’s Moral Theory, 165).

Kant thinks that the three formulations issue extensionally equivalent sets of practical judgments. But there isn’t any doubt that Kant thinks that this is no accident and that the three formulations are related more intimately yet than this. As he puts it, “each of them unites the others in itself” (436). So, Kant regards the Categorical Imperative, in its various formulations, as having some type of important internal unity. His language suggests that he thinks the three formulations are related analytically. As I’ll discuss in a moment, they don’t seem to be.

The second and third formulations, Kant says, are useful for moral education and for general moral deliberation because, as a contingent matter, they are more likely “to gain a hearing for the Moral Law” from most people (437). It seems people were put off, in Kant’s day as in ours, by any hint of abstraction.
According to the first formulation, the Formula of the Universal Law, the Categorical Imperative says ‘Act only according to that maxim through which you can at the same time will that it should become a universal law’ (421). Kant has a legislative conception of practical reason; he thinks that an agent’s voluntary or intentional behavior rationally commits him to various principles of choice, maxims. This formulation requires that we act upon maxims that are consistent with the various rules of rational willing Kant posits. The second formulation, the Formula of the End in Itself,\(^{175}\) directs our attention to those people who are affected by our behavior and requires that we never treat another human being as a mere means, but always as an end in herself (429). This is Kant’s way of saying that each person is to be treated with dignity, as something with intrinsic, objective, and absolute worth (428). The third formulation, the Formula of the Kingdom of Ends, indicates the interdependence of people within the greater human community and the complex social nature of the final moral goal, a just society. It requires that each agent act as a citizen in the kingdom of ends, a “union of different rational beings in a system of common laws” (433). This is Kant’s way of saying that, given our own personal lack of self-sufficiency, each of us should behave in a way that sustains a workable, cooperative community and is thereby respectful of the equal moral standing of each human being within society (431, 436-437).\(^{176}\) In summary, the first formulation demands volitional

---

\(^{175}\) The titles are not Kant’s, though I’ve adopted standard names for the various formulations.

\(^{176}\) For a discussion, see J.B. Schneewind ("Autonomy, Obligation, and Virtue," 310).
consistency; the second that we treat people with respect; the third that we be good citizens, fair to others.

Kant makes two important claims about the relationship between the first formulation, on the one hand, and the second and third, on the other. First, though Kant regards each of the various formulations as (having) equivalent (practical implications), he takes the first formulation—the Formula of the Universal Law—to be basic. Kant thinks that the other two formulations of the Categorical Imperative can be derived, in some way, from the Formula of the Universal Law. The direction of Kant’s argument is from reflection upon the nature of rationality and the conditions of rational agency to the universal, rational authority of the rule expressed by the first formulation, and then from the universal, rational authority of the first formulation to the universal, rational authority of the rules expressed by the second and third formulations. For instance, Kant says that the Formula of the Universal Law “will lead to” the recognition that every human being has “objective and intrinsic worth” (436, 437-438), which is the core of the Formula of the End in Itself. The first claim, then, is that ‘The second and third formulations can be derived in some way from the first.’  

177 For an argument that the essence of respect isn’t a matter of treating a person as an end in itself, see Frankfurt, “Equality and Respect”). For a response, see Joseph Raz, “On Frankfurt’s Explanation of Respect for People”).

178 There is a lengthy literature challenging this claim. For a discussion, see Pogge (“The Categorical Imperative”). For an argument that the Formula of the Universal Law and the Formula of the End in Itself generate different practical judgments, see Hare (The Moral Gap, 142-169).

179 The text of the *Groundwork* does not proceed in a linear route from reflection upon rational agency (the idea of a fully rational being) to an argument for the first formulation, and from the first formulation to an argument for the second and third formulations. Kant is, as Roger Sullivan puts it, “continually distracted” by the attempt to reveal to his readers that the Categorical Imperative is implicit within the “natural sound understanding,” common sense moral intuition (151). So, he often argues from (what he takes to be) our moral intuitions to the claim that the Categorical Imperative underlies them. This helps, Kant says, “common reason” by “drawing its attention to its own principle, in the manner of Socrates” (404). For a
The second distinguishing feature of the Formula of the Universal Law is that it indicates why the Categorical Imperative is a rule of rational willing, not merely a principle in normative ethics. Of the three formulations, the Formula of the Universal Law has the clearest license to be considered a thoroughly formal principle, the practical principle of non-contradiction. Whereas the other formulations indicate the special emphasis of respect for human dignity and for human social equality in Kant’s normative ethical theory—the “matter” of the Moral Law, in Kant’s language—the Formula of the Universal Law expresses, tersely, the principle of volitional consistency the violation of which Kant regards as a type of practical irrationality. The second claim, then, is that ‘The second and third formulations contain matter but the first only form’.

Whether Kant is able to make good on both of these claims is of central importance to his argument. Is the substance of the Moral Law—the demand to be respectful and fair—in some way derivable from the logic of agency?

At first glance, the answer seems to be ‘no’. Kant’s two claims pull in different directions. The first declares a strong relationship between the various formulations: derivability; the second a serious distinction: one contains only form, the others matter. Critics often remark that the second and third formulations cannot be logically equivalent to the first, given that—according to Kant’s own description—the second and third include matter and the first only form. But there is a question about how best to interpret Kant. A straightforward reading of several passages could easily lead the reader to think that Kant regards the relationship between the various formulations as thoroughly analytic (425, 436), perhaps he thinks, instead, that “outside” propositions are to be included in the derivations. This is a question—partly straight discussion of the actual path Kant’s chain of argument takes in the *Groundwork*, see Paul Guyer (“Introduction,” xi-xliv).
interpretation, partly a matter of charity in interpretation—that any full discussion of Kant’s theory and argument would have to address.\footnote{One relevant issue is whether Kant does include, and in a legitimate way, aspects of his “practical anthropology”—which includes empirical claims about human beings—within the derivation of the Moral Law. Or whether his appeals to practical anthropology are only a part of what are meant to be helpful illustrations of how the Moral Law does lead, as Kant thinks, to common sense moral judgments. See Guyer (“Introduction,” xi-xlv). His most direct remarks are on page xvi. Sullivan also comments: Kant “is continually distracted by his concern to explain and defend specifically human moral agency” (Immanuel Kant’s Moral Theory, 151). Another relevant issue is the status of Kant’s claim that your exercise of agency rationally commits you to valuing your capacity for reason, which makes rational agency possible. Is this claim itself “contained” within the concept of a law? Is it a rule of rational willing but not a condition for lawfulness?}

I don’t intend to settle this question. In my reconstruction of Kant’s argument, I will include in the derivations various aspects of Kant’s concept of a law and of Kant’s practical anthropology, as well as a transcendental principle. But I will leave it an open question whether each of the propositions I include are in some sense “contained” in the first formulation, or whether any are added from “outside” of it.\footnote{It seems Kant would claim that the aspects of the concept of a law are “contained” within the first formulation in that the concept of a law is mentioned in it. It’s not clear whether the same could be said of the various claims about “practical anthropology” made in Kant’s discussion of the derivation. As mentioned, it is also not clear whether the transcendental principle Kant appeals to is “contained” in the first formulation. I will address these issues in the fourth set of objections I raise against Kant.} My concern is whether the derivations are sound, not whether they are thoroughly analytic.\footnote{Of course, if the relationship were analytic, and Kant could establish the first formulation, then his argument for the second and third would go through. I am willing, though, to consider whether it might go through even if the relationship isn’t plausibly construed as thoroughly analytic. As I’ll discuss, some commentators defend Kant by adding premises to his argument. The question is then whether these premises are true, and whether they are consistent with Kant’s stated goals, such as to reveal that the Moral Law is valid for any rational being.} Is there some argument, which appeals to propositions that are all true, or at least sufficiently probable, that gets us validly from legitimate formal rules of rational agency to the substance of the Moral Law? That’s the question I aim to address.
In summary, Kant’s primary tasks are (1) to indicate how these three formulations of the
Categorical Imperative, at a glance seemingly so distinct, are in reality deeply unified—so
much so that he thinks we ought to consider them different formulations of one and the same law
—and (2) to reveal how they are each valid for any rational agent as such.

Kant’s overarching vision for human life is that living morally is tantamount to a life of
rationality, personal freedom, and human dignity. Encouraging his reader to (greater) moral
conviction, Kant remarks that we would deserve the esteem of other people, and in an unqualified
sense, if only we were to exercise our will in a consistently dutiful manner. A good will “would
sparkle like a jewel in its own right” (394), and is the only thing “in the world” or “even beyond
the world” that is good in itself (392-393). Each of us has dignity—moral value—simply in
virtue of having an ability to reason; even a person who is not particularly morally good has
objective worth simply in virtue of his humanity. But if only we were to exercise our faculty of
practical reason properly—and so, be morally good—we would take on a special type of dignity,
which our rationality, the central aspect of our human nature, makes possible for us: we would be
good specimens of humanity, an especially estimable thing to be. (Or, we could say, good
specimens of rational agency, for Kant it comes to the same thing.) For all the Moral Law’s
supposed severity, Kant thinks it is awe-inspiring and esteem-enhancing, too. And good, hard

---

183 I suspect that the Categorical Imperative, in its various formulations, is usefully regarded as a distillation
of an overarching vision of a meaningful human life in community. To understand the Categorical
Imperative, we need to understand this overarching vision, along with a number of assumptions Kant brings
to it. Each of the formulations attempts to capture an important aspect of this complex overarching vision.
But there isn’t any simple, relatively short description of the Categorical Imperative that captures its full
import. To understand the Categorical Imperative, at a theoretical level, you must understand the broader
conceptual scheme. The unity will surface—Kant thinks—from an examination of the larger vision, and it
manifests itself within the will and activity of a fully rational being. For a similar idea, see Pogge (“The
Categorical Imperative,” 189). Pogge says that the equivalency of the three formulations is a “prescription"
rather than a “description,” which suggests that the reader has to work her way to an understanding of how
they really do fit together.
thinking about the legitimate rules of rational agency and the character of the will reveals a surprisingly fertile and yet deeply unified set of prescriptions for practical living that includes the Moral Law, a set of prescriptions that are rationally required—obligatory—but personally dignifying.\textsuperscript{184} The Moral Law imposes, but a life of duty has profound consolations.\textsuperscript{185} 

Kant’s argument for this constellation of claims begins with the first formulation of the Categorical Imperative. For Kant thinks that a fully rational being, were she to have any maxims, would necessarily employ the Formula of the Universal Law in her practical deliberation (412, 434),\textsuperscript{186} and it is from this formulation that Kant attempts to derive the matter of the Moral Law. With respect to my project, we could see Kant as having proposed three tests of reflection for desire, a formal test for maxims and two substantive moral tests. The substantive moral tests, Kant thinks, forbid treating as reason-giving desires that incline an agent to behave disrespectfully or unfairly. The formal test is supposed to give the same results and indicate, in the process, why agents who act disrespectfully or unfairly are guilty of practical irrationality. I am fully sanguine that personal moral commitments grounded in, say, a desire to be respectful and fair could include norms of rational impotence. My skepticism is directed only at the claim

\textsuperscript{184} Kant operates as Socrates does: the fundamental principle of morality is already present, implicit in our every day practices, and needs the aid of a midwife to bring it to light. He is a rationalistic Socrates, as he believes that the true character of practical reason itself is being discovered in the process. Metaphysics, even the metaphysics of morals, truly practiced, is the unfolding self-revelation of reason itself. Strikingly, Kant thinks that knowledge, about which we can be certain, is available in the practical realm, though not in the theoretical realm.

\textsuperscript{185} This is a vestige of Christianity, which clearly influenced Kant’s thinking. God’s law, to sinful creatures, is experienced as an imposition; but fully redeemed humanity, for whom God’s law has become a cherished part of the heart, no longer feels this sense of imposition.

\textsuperscript{186} Kant doesn’t assume that a fully rational being, by necessity, has any inclinations, and inclinations are the source of maxims. This is one sense, but only one, in which ‘the fully rational being’ is “only an ideal.”
that any of these three tests is universally valid and at the claim that substantive practical advice, such as norms of rational impotence, can be derived from the first formulation.

The first task is to describe the first formula; the following two questions will guide our discussion: (1) Why think the Formula of the Universal Law is a rule of rational willing?, and (2) How are the second and third formulations to be derived from it? I don’t intend to give a complete answer to (2), but I will say enough to make clear both why Kant’s theory of practical reason is subject to several serious objections and how Korsgaard attempts to overcome them.

The First Formulation: the Formula of the Universal Law

Kant provides several descriptions of the Formula of the Universal Law. Here’s one:

‘Act only according to that maxim through which you can at the same time will that it should become a universal law’ (421).\(^{187}\)

There are several trip-falls on the road to clarifying this principle. First, Kant’s language is not commonplace. Anyone new to Kant is bound to wonder what a maxim is, and what it in the world it means to will such a thing as a universal law.\(^{188}\) Second, as Herman has pointed out, Kant has a non-standard theory of agency (“Bootstrapping,” 254-258).\(^{189}\) Kant rejects any version of the “belief-desire” or “belief-intention” model of agency in favor of a theory that “the

\(^{187}\) Kant also states, alternatively, “Maxims should be chosen as though they should hold as universal laws of nature” (421); or, alternatively, ‘I ought never to act except in such a way that I can also will that my maxim should become a universal law’ (429).

\(^{188}\) I don’t intend this as a criticism of Kant. If his theory is genuinely novel, as I think it is, he is compelled either to give old terms new meanings or to generate new terms.

\(^{189}\) See also Schneewind (“Autonomy, Obligation, and Virtue,” 326-327).
will is a cause” that “operates according to principles.”\textsuperscript{190} The will brings about action, Kant thinks, not directly according to desires or intentions, but through the “representation” of a principle of choice “as a law.”\textsuperscript{191} To discuss Kant, then, is to become entangled in difficult discussions concerning the fundamental elements of action-theory and agent-causation.\textsuperscript{192}

Some theorists, Herman among them, have said that Kant’s unconventionality is a virtue. This issue is directly relevant to the dialectic in my introductory chapter. Frankfurt has a subjectivist theory of the will according to which an agent’s practical point of view—his will—is defined, centrally, in terms of what the agent cares about. Frankfurt takes caring about to be a complex motivational state. This theory of the will makes it difficult, at least initially, to see how the will, so construed, has the capacity to “rule out” a desire. How do desires, however complex, disenfranchise (other) desires? Recall that I imagined a critic who claims that “at least Kant provides an account of the will that makes sense of the idea that it can disenfranchise desires.” Broadly speaking, Kant’s theory makes two moves that could help to make sense of Frankfurt’s intuitions. First, Kant claims that a desire is associated with a principle of choice, a maxim; second, he includes in the will a test of consistency for maxims. I’ve already argued, in chapters 2 and 3, that there are resources to account for these intuitions available to subjectivists. In what

\textsuperscript{190} Herman thereby criticizes Kant commentators who attribute to him the claim that a “motive of duty” is necessary to explain why a person has acted morally. Behavior with true moral worth is, according to her reading of Kant, brought about directly by the will through a representation of a maxim as a law, not by a desire. A motive, construed as an inclination, is not necessary for this process to eventuate in action, the only attitude necessary is respect. And respect for the law, Kant says, is not a motive, at least not in the sense of being an inclination (401).

\textsuperscript{191} Kant contrasts acting from inclination and from duty. Dutiful behavior is brought about by respect—“respect for the law”—and Kant does not regard respect as an inclination (400-401). For an objection to the claim that action can be brought about without an inclination, see Dreier (“Humean Doubts”).

\textsuperscript{192} Any benefits that might happen to come from Kant’s theory, Herman admits, will have to be weighed against the costs of the metaphysics that come with it.
follows, I’ll argue that Kant’s non-standard theory is problematic. He’s right neither about the claim about the connection between desires and maxims, nor about the validity of the test for maxims.

There is a third trip-fall. It is clear that Kant regards the first formulation as a principle of volitional consistency. (He calls it the practical principle of non-contradiction.) But there are different types of consistency relevant to the activity of willing. The following principles are examples:

1. ‘Make similar practical judgments in similar cases’, which is a matter of consistency in judgment,
2. ‘Make practical commitments that are mutually consistent’, which is a matter of consistency among the elements within your practical point of view—“synchronic” consistency,
3. ‘Live up to whatever practical commitments you presently have’, which is a matter of consistency between your practical point of view and your behavior—“prescriptivity,” and
4. ‘Have stable practical commitments over the course of time’, which is a matter of maintaining a uniform practical point of view across time—“diachronic” consistency.

Which types of consistency does the Formula of the Universal Law include? Any or all of these four? Any others?

---

193 By ‘practical commitment’ I mean any of a variety of things such as an intention, a plan of action, an action-guiding ideal, or a principle of choice. If Kant were to articulate these principles, he would, of course, put them in terms of principles of choice.

194 I borrow “synchronic” and “diachronic coherence” from Frankfurt (“Taking Ourselves Seriously,” 19).

195 This first principle is not a principle of volitional consistency, but it is a type of consistency relevant to the exercise of the will. We are able to live up to our practical commitments in the fullest sense only if we are able to apply them accurately to the choice situations we confront.
As I see it, the Formula of the Universal Law is strikingly complex, and each type of consistency plays a role in the thought experiment at the heart of the formulation. When Kant has us conceive of a maxim as a universal law, we are to imagine that every person in the human community exemplifies each of these types of consistency. The central question is ‘What if every person in the human community both self-legislated the maxim (presently being evaluated by the Categorical Imperative) as a part of his practical point of view and thereafter judged and behaved consistently in accordance with it for the duration of her life?’ Kant would have us ask whether such a scenario is possible, and whether, if it is, the agent would be able to act in accordance with this maxim “at the same time.”

Once Kant’s complex, revolutionary ideas are relatively clear, it will be time to turn to justificatory questions. Why is the type of inconsistency forbidden by the first formulation of the Categorical Imperative a vice? Why is it a form of practical irrationality? (Consistency isn’t always a virtue of practical rationality. Willfully sticking to a business venture falling precipitously into ruin could be stupidly stubborn rather than admirably consistent.)

---

196 The principles are each, according to Kant, rules of rationality, and Kant permits us, as the third formulation says, to self-legislate a maxim only if it could be a rule for a community of rational beings. In this remark, the term ‘rational being’ refers not merely to a being with the capacity for rationality, but a being that is exercising his rationality in a proper manner. The maxims you self-legislate have to be capable of being rules that a society of fully rational agents could follow.

197 And, Kant himself wonders, is moral evil “of a piece” with practical irrationality? Also, what grounds the practical authority of the rules of volitional consistency? Is it reason itself? It seems to me that these principles of consistency gain much of their normative authority from the role that consistency plays in helping an agent secure his various subjective goals, i.e. they have subjectivist groundings. Harry Frankfurt has pointed out, for instance, that human beings have a strong practical reason to stave off boredom, and so to have useful work. A lack of diachronic coherence threatens this goal. As Kierkegaard argues in Either/Or, significant enjoyment requires a sustained interest in something. Do they have additional, non-subjectivist groundings? Are they basic, the principles we begin from, but don’t reason to?
The immediate goal is to clarify the Formula of the Universal Law and Kant’s derivations of the second and third formulations from it. Later in the chapter, I turn to the justificatory questions, largely by considering four sets of objections to Kant’s argument.

What is a Maxim?

The concept of a ‘maxim’ has a central position in Kant’s theory of agency and in his statement of the Categorical Imperative itself. First, a maxim is a principle of action, and according to Kant’s theory of agency, rational and free action is done from a principle, not from a desire. In Kant’s view, central to our capacity for freedom is our ability not to be pushed around by desires, but instead to self-legislate and then act upon principles of choice; central to our capacity for rationality is our ability to evaluate principles of choice to see whether willing them would have genuine objective worth. Second, a maxim is the proper input into the test of the Categorical Imperative. It is the maxim of an action that is to be tested by the Categorical Imperative, to see whether it can be willed, and so acted upon, with proper consistency.

Maxim-Delineation and Maxim-Specification

To test the maxim of an action, it is necessary to determine what it is. How do we get from a clamorous desire and its claim ‘Do this. Do this’ to the maxim? It would be useful to have a clear account of what a maxim is and a set of practical guidelines for how to specify them in real life cases. Kant doesn’t supply either. Despite its importance, Kant does not systematically explicate the concept of a maxim. Serious commentators have delineated it in importantly

---

198 See T.C. Williams (The Concept of the Categorical Imperative, 13-21).
different ways, and this is not the place to indicate what each has said. I will follow, at a very general level, Thomas Pogge’s interpretation, a “maxim [M] is . . . an ordered triplet consisting of a type of circumstances S, a type of conduct C, and a type of material end E: M = <S, C, E>” (“The Categorical Imperative,” 189-190). This tripartite definition tracks what Kant himself says. But it doesn’t answer several important questions, which I will discuss in due course.

As for maxim-specification, Kant provides only the most rudimentary practical guidance. In part, this is because he’s not very confident in our ability to do it accurately. It’s an empirical matter, and so our judgments will be no better than probable; and our inclinations commonly mislead our introspective judgments: we often misjudge our own maxims out of self-love, wanting to “flatter ourselves with a pretended better motive.” On this point, Kant’s

---

199 I won’t be following Sullivan’s interpretation. He claims that many of our everyday practical decisions are not directly subject to the Categorical Imperative. The Categorical Imperative governs which maxims are permissible, and maxims are highly general “life policies.” These life policies govern our day to day practical decisions. Korsgaard might be seen as advocating a version of Sullivan’s view, but she does not seem to think it is Kant’s view. One problem with Sullivan’s view is that, if it happens to be a correct account of Kant’s theory, it is hard to see how Kant could continue to claim that we can catch an agent in a violation of the Categorical Imperative for any instance of misbehavior. If my day to day decisions are not subject to the test of the Categorical Imperative, these decisions, even if morally objectionable, can’t be forbidden by it.

200 My interpretation of Kant is at least consistent with most of Korsgaard’s remarks about the concept in Sources. But, in my view, she doesn’t present a consistent account. Her appeal to the concept of a practical identity makes it unnecessary for her to have a theory of maxims. A practical identity contains substantive rules for behavior and volition and is not a formal rule for the evaluation of principles of action. This helps Korsgaard avoid the empty formalism objection. Even so, Korsgaard continues to speak, unhelpfully, of maxims.

201 I suspect there is another explanation: he is focused on a different question. Kant’s stated aim in the Groundwork is to establish the validity of the Categorical Imperative. A close reading of the text does not establish what I am about to say, but I suspect that Kant, in the examples he discusses, simply stipulates a maxim and spends his time and energy talking about how the first formulation would evaluate it. He is saying, in effect, “Suppose the maxim of the bad Samaritan’s action were such and such; what would the test of the Categorical Imperative say about it?”

202 This is why Kant thinks that it is easier to identify what our duty is—what maxims we are forbidden, permitted, and obligated to act upon—than to know whether we are actually fulfilling it. The Moral Law, he thinks, is more easily apprehended than the human heart.
skepticism runs very deep. There is the chance that our introspections are always mistaken, and Kant more than entertains the possibility: “a cool observer . . . [is bound] to be doubtful sometimes whether true virtue can really be found anywhere in the world.” But the salient point—the one that is relevant to my evaluation of Kant’s theory of practical reason—is that Kant is committed to the thesis that there is a fact of the matter with respect to maxims: each action is done from a maxim. (Call this ‘weak universalism’.) Kant thinks that it is the genuine maxim that must pass the test of the Categorical Imperative, and my focus is the legitimacy of the test, not with any view about the frequency with which human behavior displays genuine moral worth. If there happen to be serious grounds for skepticism about the truth of weak universalism, then Kant’s theory of practical reason itself is threatened. My suspicion is that Kant fails to explicate the concept of a maxim and give adequate practical advice about how to specify them because it can’t be done without leading to serious problems for his general theory. This, in any case, is what I’ll argue.

What Does Kant Say?

Let’s begin with what is clear from Kant’s remarks. Kant says that a maxim is “the subjective principle” of an action (400-401, 421). The principle of an action is the general rule or

---

203 For remarks such as these, Kant has a reputation for uncompromising moral austerity. One might wonder: if we are not sure what our motives are, how do we know that we are not sometimes morally better than we suppose? Why not: “A cool observer is bound to be doubtful sometimes whether true vice can really be found anywhere in the world?” I agree with Harry Frankfurt’s assessment, “It must be said . . . [Kant] does not give the impression of being either harshly indifferent to ordinary human feelings or unsympathetic to familiar aspects of human weakness. Indeed, there is something poignant and rather sweet in his sorrowful allusions to the inherent frailties of human character and to the anxious maneuvers of self-deception in which we attempt to conceal them.” See Frankfurt (Reasons of Love, 76-77). But whether Kant is unduly circumspect about human moral capacity is not my concern.
policy from which the action is genuinely done. Kant thinks that each action—each instance of voluntary or intentional behavior—is done from a principle or policy that can be formulated as a rule and stated as an imperative. To say that a maxim is the subjective principle of an action is simply to say that the maxim is the prescription made by the subject who engages in the behavior. It is operating within his subjective psychology, as an impetus to action.

Kant distinguishes between “subjective” principles and “objective” principles, which he calls “imperatives.” An objective principle is a rule that is rationally inescapable, “valid for every rational being” (Lecture on Ethics, 43). It is a principle by which any agent, as such, ought to govern his behavior. A subjective principle, on the other hand, is simply a rule or policy for action upon which, as a matter of fact, some particular agent is thinking about acting. The question for a maxim—a subjective principle, being considered by an agent—is whether it ought to be followed by the agent, whether it is able—because it passes the relevant objective tests—to be a part of his personal will, a practical law for him. The point of the Categorical Imperative is to put an agent’s subjective principles to an objective test of practical reason.

The terms ‘principle’, ‘plan’, and ‘rule’ have all been used to define what a maxim is. Each has its own connotations, but I am not intending to be sensitive to each connotation in this discussion.

This explains why Kant chooses to call the supreme principle of morality, which he regards as valid for all rational beings, the Categorical Imperative. (In my terminology, an imperative—a universally valid principle—is a part of the rational aspect of each agent’s will.)

As Kant puts it, an objective principle “would serve all rational beings also subjectively as a practical principle if reason had full power over the faculty of desire” (401).

It goes without saying that Kant thinks that some maxims are morally and rationally permissible, and some are not. He says, for instance, “Every immoral man has his maxims” (Lecture on Ethics, 43).
Kant clearly conceives of a maxim as a private rule or a command to the self; a maxim is a *self-legislated* principle of action. Kant does not, however, consider every command to the self a maxim. Kant conceives of the maxim of an action as a thoroughly general policy: in acting, we are choosing a *type* of action, from a *type* of motive, in a *type* of circumstance. (Hence, Pogge’s definition.) Kant therefore thinks that the maxim of an action makes no singular reference. For example, if I am considering giving my pleasant, but bumbling student, Effie, a passing grade despite very poor academic work, I might say to myself, “Regan, don’t give Effie a failing grade.” Though in a clear sense this is a self-command, it is not the maxim of the action I am considering. A maxim makes no singular reference, and so my maxim describes each of us (if it mentions us at all) in entirely general terms. The maxim of an action is the general policy, making no singular reference, under which the agent is considering an action.

Maxims need to be discovered, specified through a process of reflection upon self and circumstance. This is true for at least two reasons. First, because of the opacity of human psychology, we must find a way to glimpse, by means of the most incisive psychological insight we can muster, through the blooming, buzzing confusion of our subjective psychology, and past

---

208 A maxim is not, according to Kant, the same thing as a conscious intention. An agent might be mistaken about the present maxim of his action—Kant thinks we often are—whereas he cannot be mistaken about a present conscious intention.

209 Perhaps, it is an implication of a maxim, applied to my circumstances.

210 Say that someone were to object to my claim that a maxim lacks singular reference. (Kant does give examples in which singular reference is made. The question, then, is whether he is speaking loosely.) The implication would simply be that the generation of a universal law, from a maxim, would include another step: stripping away singular reference. That is, the step of stripping away singular reference is necessary, and the disagreement between my objector and me would simply be about whether this step is a part of the process of maxim-specification or the process of universal-law-generation. In my view, Kant’s example is not the maxim itself; it is the maxim, applied to himself: ‘In circumstances C, with motive M, do action A’ and ‘I am in circumstances C, with motive M’.

211 As Kant puts it, “all maxims have a form, which consists in universality” (436).
whatever veils of self-deceit we happen to have put up. Second, because of a quirk in Kant’s
time theory of action. Kant thinks that an agent’s motives and beliefs “condition” but do not
“determine” behavior. The will is the cause or determinant of action, and the will operates
through thoroughly general principles, by “representing” them “as laws.” It seems more than
plausible to think that the psychological states that “condition” human behavior—that is, that
incline an agent to action—often make singular reference. I might pass Effie without thinking of
her under any general title such as ‘a student’ or ‘a pleasant person’. Thus, the thoroughly
general principle—the entity that has the proper formal structure to be put through the test of the
Categorical Imperative—has to be derived, in some way, from these psychological states, which
are not thoroughly general. What, then, is the proper way to do so?

For the sake of convenience, let’s call the psychological states that are genuinely
inclinings the agent to act—the agent’s motives and his perception of his circumstances, for
instance—the “precipitating conditions” for his action. In my grading example, my belief that
Effie is pleasant and my subsequent desire not to fail her would be among the precipitating
conditions of the action I am considering. We might then ask the question in the following way:
‘How do we get from my precipitating conditions—a desire to pass Effie because she is
pleasant—to the properly specified maxim of the action I am (considering) engaging in?’

---


213 The precipitating conditions are the psychological states that would appear in a true description of an
agent’s “motivating reason,” to use a term from Michael Smith (Moral Problem, 92-133).

214 I accept the claim that beliefs and desires commonly play a joint role in inclining agents to act. I mean
to be agnostic, at this stage, about whether there are, in some cases, other types of mental state that play a
role in inclining agents to act.

215 If there is a gap between my motivating reason and the maxim of my action, then how is it that the
maxim is necessarily a part of my subjective psychology?
Kant doesn’t really tell us. There is an analogy, Kant thinks, between the process of specifying a scientific law from particular cases and the process of specifying a maxim from a particular instance of action. In both cases, you are to proceed from the particular case to the general principle by stripping away any singular reference and then generalizing. The analogy is suggestive. In specifying a scientific law, we are to include in the law only those elements of the circumstances that played (let’s call it) an “authentic” role in the causal event: if a red ball is thrown and breaks a window, its redness is not mentioned in the specification of the relevant scientific law, but its weight and velocity are. That’s an irrelevant detail. Likewise, Kant seems to be suggesting, we are to include in a maxim only those elements of the agent’s psychology that played an authentic role in inclining him to action.\(^\text{216}\)

The Elements of a Maxim

The first question I want to address is what elements a properly-specified maxim includes. This question is a theoretical question about how to delineate the concept of a maxim, but it is relevant to the practice of maxim-specification. Without a clear target, how do we know what to shoot for?

I’ve already provided a general answer to the theoretical question. Recall Pogge’s tripartite analysis: a maxim includes an action-type, a motive-type, and a circumstance-type. As I see it, the first two elements always belong in a properly specified maxim. The trickier question

\(^{216}\) Presumably, agents do make some practical decisions directly on the basis of principles that lack singular reference. The principle ‘Never lie’ makes no singular reference, and presumably Kant acted from it. (I’m not as circumspect and suspicious as Kant.) If so, no stripping away singular reference was necessary. But we do not always engage in practical reasoning directly from thoroughly general principles of action. Our psychology includes mental states that both play a role in inclining us to act and are fixated upon particulars. In such instances, it is necessary to derive the thoroughly general maxim of my action from my attitudes, which I expressed within my practical deliberation at least partly in particular terms.
is whether the third—circumstance-type—does. A properly specified maxim includes an action-type because a maxim is a principle of action. The purpose of a maxim, its function with respect to practical reason, is to tell you what (not) to do, how (not) to project yourself into the course of events, to affect states of affairs. Accordingly, Kant’s examples always include mention of an action-type: ‘make a lying promise’ and ‘deny assistance’, for instance.\footnote{And so, Kant accepts a principle of prescriptivity: the direction of fit between maxims and behavior is that, at least generally, an agent’s behavior ought to accord with the agent’s maxims. Or, as Korsgaard puts it, the “acting self” ought to obey the “legislating self.”}

It is also clear, I think, that Kant believes that a motive-type ought to be mentioned.\footnote{Kant himself does not always include the material end or motive when mentioning a maxim in his examples. For instance, in his argument for the duty of beneficence, at times he describes a bad samaritan’s maxim as ‘Deny assistance to those in need’. But perhaps Kant is speaking loosely at this point, and he would have included an end or motive if asked to describe the maxim in full. A few lines later he does attribute motives to the bad samaritan: “prudence” and “a desire for money.” Also, it is Kant’s view that rational action is aimed at an end.} Kant includes as part of his theory of rational agency the claim that action is by its very nature purposive. We cannot act rationally without an aim in view (387). Before we begin to think about how to act, we first must have a goal toward which to strive, an end.\footnote{Korsgaard also makes this point (107-108).} Also, the Categorical Imperative would not be able to generate the outputs Kant argues for if a maxim does not include reference to a motive. It is a central feature of Kant’s general theory of practical reason that an action has genuine (moral) worth only if it is done from the right motive. Not only must our overt behavior be minimally consistent with the behavioral standards—‘Don’t lie’, ‘Don’t commit suicide’—established by the Moral Law, our behavior has to be done out of respect for the Moral Law (390). Consider Kant’s example of the two shopkeepers, both of whom maintain true scales. One shopkeeper keeps true scales out of honesty, the other out of self-interest and a fear of punishment. The Categorical Imperative is supposed to distinguish
between these maxims, permitting action upon the honest and forbidding action upon the
calculating principle. But both maxims prescribe the same overt behavior: ‘maintain true scales’.
How could the Categorical Imperative discriminate between the maxims if they did not each
include a reference to motives? If their maxims made reference only to overt behavior, they
would be the same. Consequently, a properly specified maxim mentions (in general terms) both
the agent’s motive and the action it is inclining him to do. So far, so uncontroversial.

Are we to include, as Pogge suggests, the “type of circumstance” in the maxim? In one
sense, the answer is clearly yes. In Kant’s examples, we find action-descriptions that include
implicit information about an agent’s circumstances. The maxim ‘Deny assistance’ is a principle
that applies to circumstances in which someone other than the agent is in need of help; and to
describe an action as a ‘lying promise’ indicates a circumstance in which communication has
occurred. It should be no surprise that Kant’s examples generally include reference to a type of
circumstance, often implicitly. This is a common feature of the way we describe human
behavior, for often there aren’t clear lines between act-descriptions and circumstance-
descriptions. Kant often appeals to this “circumstantial information” as he imagines what the
world of the universalized maxim would be like: he points out, for instance, that a lying promise
can be made only in a world in which there is a social institution of promise-making. In
summary, since action-descriptions include reference to information about circumstances, a
maxim does, as well.

But this leaves the following issue unresolved. Some elements of your circumstances
might operate as precipitating conditions without, naturally, being a part of your action-
description. (You might not be aware of them, or their causal influence.) Say that you have

220 Korsgaard says the “form” of a maxim is the “arrangement” between motive and action (107-108).
made a lying promise to get out of financial trouble, but not out of “mere self-interest.” You are generally conscientious, but the threat of poverty has the power to make you veer from your more common scrupulousness. You wouldn’t even consider lying to get out of most other types of personal trouble, but financial problems really get to you. In this instance, your relatively fine-grained anxiety played a causal role in your behavior by eroding your resolve. A question for Kant: should we include in the properly-specified maxim a clause such as ‘in the event that you are in emotional distress’? Or should we generalize away from this aspect of your psychology and include only the motive-type, action-type, and whatever circumstantial information is built into the action-type? Kant doesn’t say.  

How Kant answers has significant implications, which I take up in the objections section.

Level of Generality

Once we determine which elements a properly specified maxim includes—an action-motive pair, or an action-motive-circumstance triplet—we must figure out at what level of generality to describe them. As I will discuss in my objections to Kant, this is not an easy question for Kant to answer.

The difficulty arises because the various precipitating conditions often admit of a range of true descriptions. Think of motives. If I am in hot pursuit of five dollars, and I am thinking of stealing it, is the motive-type to be included in my maxim the highly specific “to get five dollars,” the less specific “to get money,” or the even less specific “for personal gain”? Each description

---

221 If Kant does think that a maxim is to be derived from the agent’s precipitating conditions, it seems that the etiology of an inclination is relevant to its rational evaluation. I find this antecedently counterintuitive. My proposal implies that a rationally impotent basic desire is rendered rationally impotent because of its structural relationship with the agent’s ruling passion—it violates a norm of rational impotence that is a part of a personal ideal the agent cares about.
is, we could easily imagine, equally accurate. And the same question arises for action-descriptions and circumstance-descriptions.

Notice, there is no linear ordering of maxim-interpretations in terms of generality. Since a maxim includes three elements—an action-type, motive-type, and circumstance-type—a question arises about the proper level of generality for each. One candidate interpretation of the properly specified maxim might be more general with respect to motive-type, but less general with respect to action-type, than another candidate interpretation. Kant needs some principled way of stopping the process of generalization, for each element of a maxim.

Could Kant give us a principled way to choose between descriptions of varying levels of generality? In his examples, Kant seems to prefer greater generality to lesser. It is natural to ask ‘why?’ 222 At this stage, I am going to set these questions aside. I will address them in the section in which I discuss objections to Kant. My goal will be to indicate that serious problems arise for any attempt to answer these questions.

Summary Remarks

With respect to the delineation of the concept of a maxim, things aren’t entirely settled. Important questions aren’t answered. As I see it, that is where Kant’s remarks leave us. Since I am going to be describing the Formula of the Universal Law and it is a test for maxims, for the sake of illustration, it will be useful for me to have maxims to put through the test. Since Kant

222 The question I raised about precipitating conditions—is it ever the case that we ought to generalize away from a precipitating condition?—is also an issue of generality. Does, for instance, the properly specified maxim of your lying promise, when you are provoked by emotional distress to make it, include reference to distress? Or does it leave out this feature of your circumstances? As mentioned, a principle that mentions the distress is, in a clear sense, less general than one that does not; it applies to a narrower range of choice situations than the one that leaves it out.
hasn’t given us a clear account of maxims, I will either adopt Kant’s own descriptions of the
maxims he discusses, or—if Kant’s treatment of a maxim gets a little slippery, and sometimes it
does—I’ll make note of this and stipulate a maxim.

Testing Maxims for Volitional Consistency

A maxim, once properly specified, is to be put through the test of the Categorical
Imperative, the Formula of the Universal Law. Following common practice, and for the sake of
convenience, I am going to call the test contained in this formula “the universal law test.” So,
what is this test, and how does it work?

To be clear, the universal law test works in tandem with another rule of rational willing.
Kant does not regard the Categorical Imperative as the only such rule. The Hypothetical
Imperative is also an objective principle, valid for every rational agent, as such. We might take
Schneewind’s helpful advice, and see Kant as encouraging a two step process for the testing of a
maxim. (The order of the steps doesn’t matter; the only important point is that a maxim has to
pass both tests to generate a practical reason for its agent.) A maxim is to be put through the
Hypothetical Imperative to determine whether, in the agent’s present circumstances, it is
instrumentally rational for her to do what the maxim prescribes. If the maxim happens to pass
this instrumental test—doing what the maxim says would actually serve her desired ends—then
Kant regards it as rational for the agent to will the maxim “as a means.”223 The maxim is then to
be put through the test of the Categorical Imperative—the universal law test of the first

---

223 Kant’s way of thinking about instrumental rationality is complicated by his commitment to the view that
action is done from principles and not from desires. Practical reasons are not communicated directly from
(rationally potent) ends to (effective) means. If doing what the maxim prescribes is instrumentally
effective, then it becomes instrumentally rational for the agent to will the maxim. This seems to me to be a
strange way of talking. It would seem that it is the action that is willed as a means, not the maxim.
formulation—to determine whether it is rational, all things considered, to will and then to act upon the maxim.

If we choose to describe things in Schneewind’s way, we must be careful, though, to point out that the second step is highly complex, and the Hypothetical Imperative (or at least some type of instrumental reasoning) enters the process a second time, within the universal law test itself. As we will see, the universal law test requires us to examine whether willing a maxim as a means in a world in which it is a universal law would be instrumentally successful. Therefore, there is an application of an instrumental principle within the universal law test itself.

A Bit of Hope for Kant’s Theory

As I see it, the Hypothetical and Categorical Imperatives generate outputs—practical judgments about how (not) to behave—only if input is provided. And there is input—a maxim—only if a creature has set herself to pursue ends. To will an end means not merely to set a positive value upon it, but to aim to pursue it. (I might see value in a medical career, but choose a different path.) In Kant’s view, if a creature does not will any ends, if there is nothing that she sets herself to do or to pursue, then there is nothing for the Hypothetical Imperative or for the Categorical Imperative to test. As O’Neill remarks, the Categorical Imperative “does not give an agent who has no plans any new ones” (“Consistency in Action”).

This presents, I suppose, a way to get out from under the stringent demands of the Moral Law. Think of the “pure wanton”: a creature with the capacity—the “mental hardware”—to will an end, who has not put it to any use. The pure wanton has various desires, the capacity for

---

224 This concept of a pure wanton is similar to Frankfurt’s. Frankfurt construes a wanton as someone who has not taken a stance with respect to his first-order desires. He does not participate in the process leading
reflection, and whatever else is necessary to exercise agency, but she simply doesn’t pursue anything. She’s not a practicing agent. According to Kant’s voluntarism about practical reasons, it follows that there is nothing the pure wanton has practical reason to do. In Kant’s eyes, willing an end, any end whatsoever, implicates you in the logic of agency: the rules of rational willing and rational agency. But she hasn’t; she’s come upon an escape clause.

Not a very attractive one, though. Surely, if Kant’s logic of agency happens to turn out correct (a big ‘if’, I think), then all of us are subject to the rules Kant advances. We aren’t pure wantons; we take ourselves and our lives seriously. We do care about how things go and pursue some of the ends we cherish. I don’t want to exaggerate; there are moments when we come off the tracks and do not pursue our normal goals. We lose heart, or suffer the psychological equivalent of not being able to find even a smidgen of motivation to get out of bed in the morning. But for most of us these moments are generally short and episodic—flashes of wantonness. And even when we lose heart about one pursuit, we usually still feel plenty of motivation to get food as soon as we’re hungry. We don’t lose all of our aims.

Could someone decide that he wanted so badly to be free of the strictures of agency that it would be sensible for him to choose to escape into pure wantonism? The costs of renouncing agency are unusually high: a life without any pursuits. Such an existence seems to lack the coherence and purpose implicit in living out a human life. It would be a meaningless existence.

---

225 This also supports the judgment that Kant’s theory is constructivist—practical reasons derive from the activity of willing. If there were no agents, no creatures setting themselves to pursue ends, capable of rational action, the world, so to speak, would be without practical reasons. (Contrary, it seems, to some types of Platonic realism, which posit the independence of the Good from human agents.)

226 See Nagel (“The Absurd”).
Meaning comes from pursuing. Not only that: Kantians sometimes press for more. Renouncing agency, they would say, is itself ruled out on grounds of inconsistency. Pursuing wantonness—attempting to escape the demands of agency—is itself a pursuing and so implicates the pursuer in a volitional inconsistency. Korsgaard, by the way, will make much of this chain of reasoning.

Kant might attempt to score a point against some of his critics here. He might argue that his theory is not insensitive to the heart’s desire. A person’s subjective concerns are represented in the context of practical deliberation, but only “through maxims.” An agent provides, as it were, the input, and the input has substantive content. Therefore, Kant’s ultimate conclusion is that whatever ends we pursue, and so whatever subjective principles (maxims) we self-legislate, once they are put through the tests, the Formula of the Universal Law will imply that we ought to treat others with respect and as equal citizens in the kingdom of ends.

Here we come to Kant’s answer to the question posed earlier: How can matter be derived from form? The material ends of an agent are represented in her maxims, and this is supposed to give hope to the defender of Kant that the matter of the Moral Law can be soundly derived from a formal principle. It is not the formal principle alone, but the formal principle along with the input that agents supply through maxims, that is to generate substantive practical judgments, including the matter of the Moral Law. It is in this way that the universal authority thesis is to be established.

---

227 For a beautiful defense of this claim, see Richard Taylor ("The Meaning of Life").

228 The implication would be that the only way to escape the rules of agency without being subject to this charge is to “fall” into pure wantonness, by accident rather than by intention.

229 In Kant’s words, “the free will stands, as it were, at the crossroads halfway between its a priori principle which is formal and its a posteriori incentive which is material” (400).
The Hypothetical Imperative

According to Kant, as a first approximation, the Hypothetical Imperative states, ‘Whoever wills an end ought to will the means’.230 As written, this is an instrumental principle.231 Kant thinks the will addresses the maxims presented to it by impulses to action, where these impulses are the consequence of at least some rudimentary instrumental thinking. The Hypothetical Imperative kicks in after a certain amount of practical deliberation, whether short or lengthy, has brought the agent, aiming at some end, to consider doing some action to bring about that end. It is the purpose or function of the Hypothetical Imperative to test a maxim to see whether it is instrumentally rational for the agent to do what it prescribes in his present circumstances.232

Our capacity for reflection makes it possible for us to ask whether acting according to a particular impulse, and its maxim, would (best) further our ends. But the Hypothetical Imperative only tells the agent to will an action “as a means”; it does not, by itself, rationally compel action. For the maxim must pass a further test for it to be rational, all things considered, to act upon it.233

---

230 Is the Hypothetical Imperative a principle of rational virtuosity or of rational advice? That is, is its purpose to correct for poorly-grounded means-end beliefs (in which case it is for rational virtue) or for mistaken means-end beliefs (in which case it is for rational advice)? Kant doesn’t say.

231 Kant often describes the Hypothetical Imperative as a principle of prudence or of happiness, which reflects his tendency to see human agents as suffering from a struggle between “the dear self” and the Moral Law. (Kant believes in a strong human tendency to egoism, to undue love of self.) But the Hypothetical Imperative applies to impulses that do not have “benefit to self” as their object. The general purpose of the Hypothetical Imperative, as stated, is to guide the agent in his pursuit of his desired ends, whatever they happen to be.

232 Does this test do anything more than simply check to see whether the action the maxim prescribes is an effective means? Then in what sense, we might ask, does the Hypothetical Imperative test the maxim?

233 Here are the qualifications Kant makes to the Hypothetical Imperative. He repeatedly contends that whoever wills an end, if he is fully rational, wills also the “necessary” or “indispensable” means to those ends “which are in his power.” The qualification ‘within his power’ reflects that Kant accepts ‘ought’ implies ‘can’. This principle implies that a human being is under a requirement of (instrumental)
The Universal Law Test

The Formula of the Universal Law appeals to the concept of a law. It says, [A] ‘Act only according to that maxim through which you can at the same time will that it should become a universal law’ (emphasis mine). At another point, Kant describes this formulation as saying, [B] “Maxims should be chosen as though they should hold as universal laws of nature.” In Kant’s view, it is a part of the very idea of a scientific law that it is universal in two ways; it (i) makes no singular reference and (ii) holds in every case. Kant takes the idea of a practical law from the idea of a scientific law (437). The question to be asked of a maxim, according to Kant, is whether it has each of the features requisite to be a law. A maxim is a valid principle of choice for an agent only if it meets these two conditions.

Since a properly specified maxim makes no singular reference, it meets, by definition, the first condition for lawfulness.234 The process of maxim-specification, if done properly, ensures this. What remains to be seen of any particular maxim is whether it is able to “hold in every case.” The thought experiment at the heart of the universal law test aims to discover whether it is “universalizable.” As I see it, there are two criteria for meeting the universalizability condition.

234 If you’re inclined to think that a maxim is able to make singular reference, then the stripping away has to be done before the thought experiment is executed.
A maxim is to be treated as a practical law—an appropriate guide to action, and a proper aspect of the agent’s personal will—only if it meets both of these two criteria. First, a maxim has to be, in a logical sense, capable of being a universal law of nature. The answer to the question “Could there be a world in which the maxim is a universal law of nature?” has to be ‘yes’. [A] and [B] both gesture at this requirement. But [A] seems to indicate an additional requirement. Even if a maxim could be, logically speaking, a universal law, it is necessary to check to see whether we could act upon it in the event that it is. We’ll have to say a bit more to make sense of what Kant means by this.

The Thought Experiment

The universal law test can be described as a thought experiment, as long as we’re careful. Kant does not take thinking through the test to be an act of mere imagination; it is work to be done by the understanding. The maxim is being examined to see whether it has the appropriate formal properties, whether it really is “law-like,” whether it holds—or is able to—in every case. And this is a matter of discovery. (The Categorical Imperative puts theoretical reason to work, for practical purposes.) But for the sake of illustration, it is helpful to think of the test as a thought experiment that an agent is to go through in making a practical decision. The best way

---

235 In my view, it’s not merely a question of whether such a world would be attractive in various ways, or whether an agent would consent to being in such a world, but whether it would be logically possible. I doubt that Kant would find a world in which no one assisted each other an attractive world, and I doubt he would consent to live in it (if he had friendlier choices), but he does not rule out the maxim ‘Deny assistance’ on these grounds, as we will see.

236 I do not mean to suggest that Kant is subject to the objection, often directed at consequentialism or utilitarianism, that it is not possible for a human being to go through this test before each action for it would leave that person with no time left for living his life. Kant might well think that an agent, from experience, comes to see which maxims tend to pass the test and which do not, and so does not need to begin the test
to tell whether a principle could possibly be a universal law of nature is to (try to) imagine it becoming one. The best way to tell whether it is possible to act upon a maxim at the same time that it is a universal law is to (try to) imagine doing so.

The first step in the thought experiment is to “universalize” the maxim. In evaluating a maxim, Kant asks an agent to (attempt to) conceive of a world in which each human being acts, as though compelled by a scientific law of nature, according to its dictates. What if everyone were to adopt, and to successfully obey, the maxim? We are to imagine that the maxim is obeyed so faithfully that it might appear, from the perspective of a third person observer, to be a scientific law. Under normal conditions, at the moment the temperature falls to zero degrees centigrade, water begins to freeze; and under normal conditions, at the point that an object of sufficient mass hits a window with sufficient velocity, the glass breaks. So, with the same law-like regularity, under the relevant conditions, the action mentioned in the maxim is done. If we construe a maxim in the sense suggested by Pogge, the maxim includes the agent’s precipitating conditions, the beliefs and motives that are inclining the agent to act, stated generally. Then we are to (attempt to) imagine a world in which everyone who finds herself under these precipitating conditions obeys the principle—and does so with “law-like” regularity.

Notice that a world in which a maxim is a universal law is a world in which agents successfully obey the four principles of volitional consistency I mentioned earlier. Kant thinks that a fully rational agent would exemplify each; and in the imagined world, the human anew each time he is making a decision. (And, of course, the consequentialist/utilitarian has the same response available.)

---

237 This thought experiment does not, as sometimes said, commit Kant to consequentialism, and so implicate him in a theoretical inconsistency, both denying and accepting consequentialism. What makes the destruction of the institution of promise-keeping problematic is that it implicates the agent in inconsistent willings. See Frankena (Ethics, 26).
community is a “system of rational beings.” If the maxim is a universal law, each person has a stable and persistent commitment to the maxim across time (“diachronic coherence”), applies the principle to his choice situations in a consistent manner, making proper practical judgments in accordance with it, and behaves accordingly. This also presumes, it seems, that the hypothetical agents do not have, or at least do not obey, principles of choice that conflict with the maxim (“synchronic coherence”). A maxim is to be tested to see what would come of things, if it

---

Kant accepts a number of principles of volitional consistency. One way to clarify Kant’s idea is to imagine, as he does, what a fully rational agent would be like. A fully rational agent, Kant thinks, constructs a practical point of view that consists of mutually consistent principles of choice, and thereafter judges and behaves in accordance with them. His life, that is, displays a remarkable unity, within his practical point of view, and between his practical point of view and his behavior. The first—‘make similar practical judgments in similar cases’—is a principle of relevant difference, and Kant clearly accepts the claim that an agent is justified in making different practical judgments in two choice situations only if these choice situations differ in a relevant way. This is to say, Kant requires—plausibly, I think—consistency in practical judgment. We ought, rationally speaking, to apply our principles to our choice situations consistently. The second principle—‘adopt mutually consistent practical commitments’—reflects Kant’s belief that an agent that fails to would suffer from consistent frustration of one or the other conflicting practical commitments. The third principle, which I will call ‘prescriptivity’, states ‘Live up to your practical commitments’. This principle is grounded in the idea that the point or purpose of adopting—that is, self-legislating—a principle of choice or a plan of action is to employ it to guide your future behavior. And so the proper “direction of fit” between an agent’s principles of action and his behavior is that a person should shape his behavior in accordance with his principles, not the other way around. (The other way around, generally speaking, is mere rationalization.) Kant regards the will as having a function, and the authority to exercise it: the will is to guide behavior. There are, of course, instances in which an agent ought to reconsider his principles of choice—Kant, of all moral philosophers, would not deny this: the Hypothetical and Categorical Imperatives are tests for the practical evaluation of principles of choice. And so, if your unerring obedience to a principle of choice consistently leads to the frustration and not the satisfaction of your desired end, you ought to re-evaluate, in light of this information, whether the Hypothetical Imperative endorses it, as you had previously thought. But, as Korsgaard puts it, Kant thinks it is the work of the “legislating self”—the will—to tell the “acting self” what to do. The will, Kant thinks, has the authority to govern action; hence, prescriptivity. This is how things would go for a fully rational agent: he would shape his behavior in accordance with his principles. The fourth principle, which I will call ‘agent-stability’, states that an agent should create and foster a stable and persisting set of practical commitments. This principle is grounded in Kant’s idea that it is not possible to be a stable and persisting agent, and to exercise agency in a principled manner, without having a stable and persisting practical point of view (consisting entirely of principles of choice) from which to make practical decisions. Incessant waffling, within the personal will, would render a person incapable of action, properly speaking. If we are to project ourselves into the course of things in a principled and not an arbitrary manner, we need stable practical commitments. Kant thinks that the question before us concerning a principle we are considering is whether we could “make it a habit” to act upon it (402). A fully rational agent has a sure and lasting foundation from which to make his practical decisions.
were obeyed in these ways by everyone. (Such a world is, Kant admits parenthetically, “certainly only an ideal” (433).)

The question, once all of this is imagined, is whether an inconsistency—Kant says a “contradiction”—arises. Kant mentions two types of inconsistency. The first is generally called a “contradiction in the conception,” the second a “contradiction in the will.” As I see it, the difference between these two types of contradiction is that a contradiction in the conception is a contradiction that arises among the various volitional states that are, as it were, directly created by acting upon the maxim in question, whereas a contradiction in the will is a contradiction between the volitional states directly created by acting upon the maxim in question, on the one hand, and a volitional state that is already, and inescapably, a part of the agent’s (personal) will, on the other. In this way, a contradiction in the conception is a contradiction “internal” to willing the very action in question, a contradiction in the will is a contradiction between willing the action and other “inescapable” aspects of the agent’s will.

Contradiction in the Conception

Commentators disagree about how to construe a contradiction in the conception. Some think that it is a logical contradiction, others practical. This is an instance in which it is not necessary for me to take sides between disputing Kant commentators. First, my neutrality is fair to Kant. If it is possible to defend his theory by arguing soundly that either a logical or a practical contradiction is both practically irrational and has serious moral implications, this would begin to vindicate Kant’s larger argument. Second, I happen to think it is natural to see Kant as thinking that the universal law test is to seek out both logical and practical contradictions. [A] intimates a logical test, [B] could be interpreted as suggesting a practical test as well. (I happen to think it is
harder to make sense of Kant’s examples if we try to see them as giving rise to logical contradictions. But I won’t press this conclusion here.)

A Logical Contradiction

Some of Kant’s language suggests that the universal law test is a logical test. Let’s call a maxim, conceived as a universal law, the “universalized counterpart” of the maxim itself. Kant says the universalized counterparts of many immoral maxims are “inconceivable” or “cannot be thought” (424), which intimates that Kant has logical impossibility in mind. But how could a logical contradiction arise from the attempt to conceive of a maxim as a universal law? Let’s consider three ways.

First, take the principle ‘At five o’clock each morning, give every person both ten dollars and nothing’. This principle is self-inconsistent. The action-type commanded by a maxim describes a logically impossible state of affairs; in fulfilling one part of the rule, you necessarily violate the other. You cannot conceive of a world in which everyone successfully obeys this rule because it is not possible for even one person to successfully obey it. If you try to imagine such a world, you will fail—not because of a lack of imagination, a failure in you, but because of a defect that could be described as internal to the maxim. A principle of action is supposed to give practical advice, this one gives practical advice that can’t possibly be followed.

Second, take the principle ‘Kill everyone else, but sustain your own life’. Logically speaking, this command could be successfully discharged by one person; it is not self-

\[\text{Perhaps it is possible for a person to attempt to obey the rule (though he would have to be especially obtuse); so, obedience to the rule could be pursued, and so willed as a universal law in that sense. But Kant has us imagine a world in which people successfully discharge the obligation, behaving as though the rule were a scientific law of nature. And this cannot happen with respect to this rule. The rule could not be a universal law of nature.}\]
inconsistent. But it is not possible for multiple people to successfully obey the command. One person’s discharging it successfully implies that another has not, even that everyone else has not. Kant’s universal law test requires that you imagine that the maxim has become a universal law: everyone obeys it each time the precipitating conditions obtain. But it cannot be a universal law of nature.240

The first two types of logical contradiction arise in attempting to conceive of a world in which a maxim is a universal law. Such a world cannot be conceived, is literally inconceivable. The thought experiment cannot be consummated—nature cannot contain contradictions—and the contradiction arises in the very attempt to imagine it as a universal law.

There is a third kind of logical contradiction that can arise. This happens if a maxim is self-referentially absurd. Consider the principle ‘Never adopt a principle of choice’. There could be a world in which nothing is ever willed—for instance, a world in which rational beings never come into existence. But a contradiction arises in the event that you will this maxim. In adopting the principle, you thereby violate it. The principle demands that you not adopt any principles of choice, but self-legislating the maxim is itself the adoption of a principle of choice: a logical contradiction.

The three principles I have mentioned each, if adopted, generate what might be described as logical contradictions. I’m not sure whether they are what Kant has in mind when he speaks of contradictions in the conception? No doubt, Kant would reject each of the candidate maxims I mentioned, and for the reasons I give. None could be, in Kant’s view, a proper part of an agent’s

240 It is possible, of course, for everyone to attempt to follow this pernicious rule, but it cannot be the case that everyone successfully obeys it.
personal will.\textsuperscript{241} But these examples certainly aren’t reminiscent of the central examples that Kant actually discusses. And to generate logical contradictions of these types from Kant’s examples, we would have to build in a number of assumptions about the imaginary world we are attempting to conceive.

Again, though, I’m not going to take an official stance. My concern is whether Kant’s argument is sound, and I don’t need to take a stance on this question to address my concern. It’s important that we get straight what types of volitional principles Kant appeals to, however he categorizes them, and how he thinks the Moral Law follows from them. I’ll simply take the universal law test to be a check list that includes the requirement that an agent check for a logical contradiction, and then for the following kind of practical contradiction.

A Practical Contradiction

Korsgaard advocates a different interpretation of ‘contradiction in the conception’. She regards it as a practical contradiction. Her interpretation focuses upon [A], in which Kant says that an action is rationally permissible only if the agent is able to will the maxim of the action as a universal law “at the same time” that he acts upon it. Since Kant thinks that, in acting, an agent is willing a maxim as a means, we might think that Kant means for us to check to see whether 

\textit{willing a maxim as a means} is consistent with \textit{willing it as a universal law}.

Clearly, the question within the universal law test is not whether willing the maxim as a means is instrumentally effective in the agent’s present circumstances—in “the real world,” as it were. That judgment is to be issued by the Hypothetical Imperative. The agent is checking the

\textsuperscript{241} As for the third type of logical contradiction, any self-referentially absurd principle would be ruled out already by the Hypothetical Imperative. So, Kant would forbid it.
maxim against the universal law test, presumably, only because it has first passed the Hypothetical Imperative and so is instrumentally rational in the real world. Instead, the agent is to project himself into the imaginary world in which the maxim is a universal law of nature and consider whether his instrumental strategy—acting upon the maxim—would be successful “there.” If it wouldn’t be, then acting upon the maxim in reality would implicate him in an inconsistency: the maxim would not be able to serve the instrumental purpose he is putting it to if everyone behaved similarly.

Korsgaard’s interpretation fits with some of what Kant says about his own examples. Take Kant’s example of the lying promise. Suppose that you are in financial straits. In order to get a loan, you ask a well-to-do friend to help you out, and you promise to repay him, with the intention not to. You would be committing yourself, Kant is assuming, to a maxim; let’s stipulate that it would read ‘To benefit yourself, make a promise with no intention of keeping it’. (Whether this is the properly specified maxim is not clear. But never mind.) Say that the Hypothetical Imperative would tell you to will to make the lying promise; it is—in your present circumstances—an effective and, let’s say, even indispensable means to the benefit you seek. For you, there is no way out of your fiscal problems without lying. So, the maxim passes the first step: you are rationally committed to willing the lying promise as a means.

Can you at the same time will this maxim as a universal law of nature? Kant says no. To see why, he asks us to imagine that every single person within the human community acts as though they are physically compelled by a natural law to make lying promises in the event that the precipitating conditions (personal difficulty and a desire to get out from under it) obtain. Kant’s idea is that the institution of promise-making would fail in a world in which this prescription is adopted and obeyed so steadfastly by all rational creatures: “no one would believe
what was promised to him, but would laugh at any such assertion as vain pretense.’’ The
institution relies upon general trust, which would be utterly eroded in the imaginary world. But in
a world where no promises are trusted, it becomes impossible for you to reap the (instrumental)
benefits of your lying promise. You’re promise would not be able to fulfill the purpose you are
putting it to, at least not in the imaginary world; it would provoke derisive chuckles, not decisive
funds.

What’s the contradiction? On Korsgaard’s reading, Kant thinks of the inconsistency as a
matter of self-defeat. You are willing both an end and its frustration. Your maxim would be self-
defeating if universalized: your action would not be instrumentally effective if everyone tried to
use it for that purpose. According to Kant—and this part I find very plausible—working at cross-
purposes with yourself is a paradigm example of practical irrationality. According to Kant—and
this part I’m not so sure about—the lying promise-maker is working at cross-purposes; she is
acting as though she both wills to sustain the institution of promise-making (to put it to
instrumental use) and to undo it (in which case it will not be of use to her). And this implies
that she is willing the contradictory ends of getting out of financial straits and of destroying her
means to.

The contradiction arises, of course, only if several assumptions are true. I want to bring
out one of these assumptions because I will be discussing its truth in the section about objections

---

242 There are several distinct interpretations of Kant’s conception of volitional inconsistency. I am
describing the interpretation that Korsgaard advocates (“Kant’s Formula of the Universal Law”).

243 Korsgaard’s interpretation does not follow this pattern: it is possible to make a promise only if someone
accepts it, and it is not possible to make a promise in a world in which no one is willing to accept one.

244 Another possible interpretation, which takes Kant to be claiming that a lying promise is logically
impossible in the imaginary world, states that the words ‘I promise’, which are performative, are not able to
create a promise in the world in which promises are not trusted.
to Kant. Kant thinks that it is legitimate to attribute to an agent who engages in voluntary or intentional behavior a variety of volitional states. First, Kant thinks action is done from a maxim, which is thereby willed “as a means”: weak universalism. Kant is also committed to what might be called ‘strong universalism’: the claim that acting upon a maxim rationally commits an agent to willing it as a universal law. To act is to reveal what type of world you are willing (i.e. consenting) to bring into existence. I suppose we might say that Kant thinks that acting for any reason whatsoever has, as it were, “eschatological implications.” It’s not that your actions will cause such a world to come into existence. (It’s a bit vain to think everyone is out to mimic you.) But your behavior indicates, Kant thinks, that you are aiming, or at least consenting, to create such a world. If you lie, you can rightly be accused of at least consenting to the creation of a world in which everyone does the same.

The social practice or institution of promise-making makes your lying promise possible. So, you are both willing to use a general institution and willing to destroy it at the same time, in the same act of willing. And Kant regards that as self-contradictory.

Contradiction in the Will

If a maxim does not generate a contradiction in the conception (of a logical or a practical variety), then it is necessary to check to see whether it generates a contradiction in the will. A

---

245 As Korsgaard herself remarks, the universal law test works best when the maxim prescribes the use a social institution, such as promise-making. It is when a social institution is universally misused that its very viability is compromised, and so a practical contradiction can be derived. It is not as clear how “natural” actions, such as harming out of vengeance, generates any such contradiction. There is no practice associated with acts of vengeance that is both instrumentally necessary and destroyed at the same time (“Kant’s Formula of the Universal Law”).

246 Is this judgment reinforced by a piece of transcendental reasoning—namely, “you can’t will something without willing what makes it possible”? In that case, if you will a lying promise, you must will whatever makes such a lie possible, and a lying promise is possible only if the practice of promise-making exists. Korsgaard’s argument will rely heavily upon transcendental reasoning such as this.
contradiction in the conception of the practical variety is a contradiction between two volitional states: willing a maxim as a means and as a universal law. A contradiction in the will is different. It has to be a contradiction between willing the maxim (as a means and/or as a universal law), on the one hand, and some other volitional state that is itself rationally inescapable, on the other. If this other volitional state were not itself rationally inescapable, the agent would be able to render his willings consistent and so act upon the maxim he is considering simply by rejecting this other, contradicting volitional state. What rationally inescapable volitions does Kant have in mind?

Take one of Kant’s examples of a contradiction in the will: a bad Samaritan act. Suppose I have some discretionary wealth, there are people who are suffering and in need of my financial help, and I walk past them in their time of need. Presumably, my motive for walking past them—assuming I noticed—is to benefit me and my own from my wealth. If we eliminate singular reference from the description of my behavior, and provide a general description of the precipitating conditions, the maxim fully specified is (again, let’s stipulate), ‘Keep more than adequate personal assets in order to benefit self and family, even in the event that others are in serious need’. Or, by Kant’s account, the maxim is ‘Deny assistance to those in need’.

Kant’s idea is that we are able to imagine a world in which the well off, in search of personal benefits, systematically deny assistance to the absolutely poor. No logical or practical contradiction arises. It is logically possible for there to be a world in which the bad Samaritan maxim is a universal law, and it is also possible to hoard your own wealth in the imagined world of the universalized maxim. (This seems to be in part because there isn’t, as in the example of the lying promise, a social institution that could be said to be both being used and destroyed simultaneously.)
Even so, Kant thinks this maxim, if willed, generates a contradiction. We are needful creatures, not fully self-sufficient. And so, even if you aren’t presently in need, you have been and will be at some point. (If you were to deny it, a Kantian might ask: Did you feed yourself as an infant? Did you build the schools that educated you? Did you grow the grain that presently sustains you? Did you build the buildings that are needed for your future job?) The Hypothetical Imperative, at various times, commands each of us to get help from others. Help from others is sometimes, in the course of a normal human life, the indispensable means to an end. And so, “the human will in so far as it is affected by nature” invariably and inescapably includes a maxim that precludes our always going on our own. If an agent wills the bad Samaritan act, he is both willing that the bad Samaritan act be a universal law and that it not be, that there be an exception to it, one that favors him. And Kant regards this as contradictory.

Summary Remarks about the Universal Law Test

My interpretation of Kant’s concept of contradictions makes clear the importance of the Hypothetical Imperative, which Kant regards as a basic rule of rational willing, in generating several of the volitional contradictions the Categorical Imperative forbids.247 Also, my account makes sense of Kant’s idea that the agent’s desires and motives, which are referenced in his maxims, provide the “matter” and the Formula of the Universal Law the “form.” A particular human being’s subjective concerns, along with facts relevant to their satisfaction, render some behavior instrumentally rational, and the first formulation of the Categorical Imperative constrains which instrumentally rational behavior is rationally permissible, all things considered.

247 Kant says that determining our duties is relatively easy (403-404). This might strike the reader as a bizarre claim, given how complex Kant’s theory is. I have no intention of defending this remark by Kant.
On my interpretation, Kant’s theory is not insensitive to an agent’s personal concerns; they do receive a hearing before the will, by way of their associated maxims. As I’ve already mentioned, this is what is supposed to give Kant and his disciples hope that it is possible to derive substantive moral implications from the application of a thoroughly formal principle. The output of a formal principle could possibly be substantive if the input is.

That said, weak and strong universalism are questionable, and there are questions about whether Kant is able to derive, from the universal law test, the content of the Moral Law.

The Derivation of the Second and Third Formulations from the First

I haven’t yet indicated how Kant derives the second and third formulations of the Categorical Imperative from the first, an important piece of unfinished business. I have revealed how Kant would employ the first formulation to scrutinize two maxims, ‘Make a lying promise’ and ‘Deny assistance’. Kant claims that these two maxims are forbidden by the universal law test, and he might point out that the Categorical Imperative—even the thoroughly formal first formulation—has substantive moral implications. But his argument, even if it were sound to this point, wouldn’t secure his claim that the second and third formulations of the Categorical Imperative can be derived from the first. The (perfect) duty ‘Never make a lying promise’ and the (imperfect) duty ‘Sometimes give assistance’ are so-called “little” categorical imperatives. The “little” categorical imperatives are supposed to be derived from each of the formulations of the “big” Categorical Imperative, which are supposed to have equivalent implications. So, they can’t be premises in the derivation of any of those formulations. Hence, we still need to figure out how Kant gets from the Formula of the Universal Law to the End in Itself and Kingdom of Ends formulations.
As mentioned earlier, it doesn’t seem as though the derivation could be thoroughly analytic. How could the mere substitution of synonyms get us from pure form to the matter of the Moral Law, the demand for respect and fairness? Substance has to enter into the argument in some way. Accordingly, Kant makes an appeal to the characteristic activity of any agent, the self-legislation of maxims. The very act of willing a maxim, any substantive maxim whatsoever, is creative of volitional states that rationally compel an agent, by way of a transcendental principle, to be respectful and fair.

The derivation has, broadly speaking, two steps. In the first, Kant applies a transcendental principle; in the second, the principle ‘Make similar judgments in similar cases’, a principle for consistency in judgment. I hesitate to say that I fully understand this argument; I find transcendental arguments difficult to track. Gerald Cohen, in his commentary about Korsgaard, speaks of his “bone-headedness . . . about the character of transcendental arguments” (186). I identify with the feeling.

The transcendental principle Kant posits is something like this: ‘You are rationally committed to value whatever makes possible the things you value’. This principle is already implicit in Kant’s reasoning about the example of the lying promise. If you make a promise, you are putting to use the institution of promise-keeping. Without the institution, he thinks, you couldn’t make the promise. So willing the promise as a means implicates you in valuing the institution that makes the promise possible.

The same move is present in Kant’s derivation of the second and third formulations. Kant thinks that agency is a matter of the pursuit of material ends, and the pursuit of material ends implicates an agent in a form of valuing. If you pursue an end, any end whatsoever, Kant would say that you “value” it. And in pursuing this end that you value, you are employing your
will, that is, your faculty of practical reasoning. Having a will—whatever mental hardware makes it up—is a “necessary ground” for the activity of pursuing. The transcendental principle implies that you are then rationally committed, in employing your will, to value it. And since (along with having an intellect) having a will is what makes you, in Kant’s technical sense, a human being, it follows, by definition, that you are rationally compelled to value your humanity. That’s the first part of Kant’s derivation. It doesn’t yet get us to the claim that you are rationally committed to respect for any human being, but we’re close.

In the second part of the derivation, Kant seems to reason in the following way. If you value your rationality, as such—or your humanity, as such—an application of the principle of consistency in judgment—’Make similar judgments in similar cases’—compels you, Kant thinks, to take the same attitude to any instance of humanity. If you value, or are rationally committed to valuing, your will, and other people engage in the same type of activity—rational activity—as you, you have to attribute the same value to them and their activity as you do to you and yours. Ergo, the second formulation. (Korsgaard glosses this: you must value humanity “wherever you find it” (120-125).) The feature that makes you a human being is no different than the feature that makes others human beings. They are, as some Kantians have said, “equally real.”

So, if there happens to be a community of such beings, if you happen to bump into some other beings with the same capacity, you are going to have to pay them the same respect you pay to your own humanity.\(^\text{248}\) As things turn out, most of us reject solipsism and admit the existence of other people with minds like ours. Kant thinks this rationally commits us to respect for them, and ultimately a willingness to work in accordance with them in a community of rational beings.

\(^\text{248}\) Again, there is a biblical source, given a rationalized, Kantian reading: “Who is my neighbor?” Kant’s biblical answer: “Anyone whose path you cross.”

The argument is sketchy, and I find the language of “valuing” slippery. Without a clear account, it’s hard to tell whether the inferences from the valuing implicit in pursuing to valuing as a necessary ground are valid. Let’s be clear, though, about two things. First, the transcendental principle has to get us to the right kind of valuing. The capacity for reflection has instrumental value. Our ability to reason often helps us to do better at satisfying our desires than we otherwise would. An instinct might incline us to behave in a way that would not get us what we want; better, then, to be able to reflect and change course. But, by this account, my will has instrumental value to me, and yours to you. Because we live in a community, I often benefit from other people’s ability to reason, too. The neighbor who puts her mind to it and figures out just how I like cookies makes batches I like better. But all this said, there are certainly people whose ability to reflect isn’t of any instrumental value to me. If the value I had to attach to my capacity for reasoning, or to yours, were only instrumental value, Kant’s argument wouldn’t seem to go through. I don’t have to value my ability to reflect, as such, and so it doesn’t seem I have to value myself, my humanity, as such.

Consequently, Kant seems to be committed to the claim that your capacity for practical reasoning has more than instrumental value to you. For lack of a better phrase, Kant seems to think you have to value it “as a necessary ground,” which, apparently, is a type of intrinsic valuing: you have to value your capacity, as such. Is Kant right that valuing something rationally compels us to value any of its necessary grounds in an intrinsic sense?

Second, I presume that Kant wants to say that you and I have a measure of personal freedom to pursue our own goals. My personal projects aren’t yours. When I self-legisllate a

---

249 This presupposes, of course, that our practical deliberation has an effect on what we do. “The will is a cause,” Korsgaard says. But Kant says we at least operate under the idea that our will has an effect upon our behavior; we operate “under the idea of freedom.”
maxim that tells me to, say, create a stamp collection, so long as the maxim passes the requisite tests I come to have practical reasons to take the necessary means. My “values” lead to practical reasons for me. Kant had better make sure that valuing another person’s humanity does not include my having to treat their personal projects as the source of practical reasons for me, at least not practical reasons as compelling as my own projects create for me. Can he? It’s hard to tell.250

I don’t happen to find Kant’s derivation, understood in this way, compelling, or even promising. In the following section, I will discuss why.

Conclusion

This concludes my logical reconstruction of Kant’s argument for the universal authority thesis and the will’s authority to categorically reject the claims of desire: the claims that fail the test of the Categorical Imperative do not deserve, Kant thinks, any voice within practical deliberation. Even if there are means to their satisfaction, they fail to play a role in generating practical reasons for the agent.

To foreshadow, Korsgaard will drop important elements of the universal law test—weak and strong universalism—from what she calls the Categorical Imperative, and she will not appeal to them in her argument for the universal authority thesis. This is one of her major departures from Kant. But she needs to posit rules of rational willing strong enough to do what Kant wants weak and strong universalism to do: compel obedience to morality and the disenfranchisement of immoral desires by revealing that our exercise of practical reason necessitates it. Her theory

250 Perhaps the transcendental principle already operates within the Formula of the Universal Law. If we think back to our earlier discussion, in Kant’s claim that you are guilty of a contradiction in the will when you will a lying promise: if you make a lying promise, you are employing, and so valuing, the institution of promise-making, and yet de-valuing it at the same time, in attempting to destroy the institution. What Kant does to get from the first to the second is to apply this rule of rational willing.
employs a transcendental principle, and the soundness of her argument turns on whether some version of such a principle is both valid and as fertile as she and Kant suggest.

Objections to Kant’s Argument for the Universal Authority Thesis

A great variety of objections have been directed at Kant’s argument for the universal authority thesis and the concepts that Kant appeals to within it. I do not intend to discuss each of them. Only four, each of which I take to be both serious and directly relevant to Korsgaard’s attempt to refurbish Kant. Korsgaard’s argument might be seen as Kant’s, adjusted to overcome these objections.

Korsgaard herself only mentions, and so explicitly aims to overcome, several versions of the objection commonly called the “empty formalism” objection. These objections concern the “fertility” of the Categorical Imperative, its capacity to generate substantive practical implications, in particular rules with moral content. To overcome these objections, Korsgaard appeals to the concept of a practical identity and its characteristic role in practical deliberation; an agent’s self-conceptions, she thinks, are the source of the substantive practical advice an agent ought to follow. I am going to discuss three more objections to Kant because I think that Korsgaard’s appeal to the concept of a practical identity could plausibly be interpreted as a response to them, as well. A fair accounting of her argument must take into account each of its strengths.

251 For example, I am not going to discuss the question ‘How is it that the violation of what we are to regard as a purely formal principle—a quasi-logical rule—amounts to a moral evil? In Kant’s own words, moral evil and formal inconsistency do not seem “of a piece.”
The various objections to Kant that I am going to discuss are not easy to disentangle. So tightly woven is Kant’s system that objections to one part often have serious implications for other parts. For the sake of clarity, it is useful to begin with the following objections, phrased as skeptical questions.

(1) Why think each action is ‘done from’ a single and thoroughly general principle of action?

(2) Why think an agent is rationally committed to willing the maxim of his action as a universal law?

(3) Why think the Formula of the Universal Law, a formal principle of volitional consistency, will imply respect and fairness?

(4) Doesn’t Kant illegitimately appeal to premises that are not merely claims about rational agency within his derivations of the matter of the Moral Law?

These four skeptical questions challenge, respectively, Kant’s attempt to generate input for the test of the Categorical Imperative, the legitimacy of the universal law test, its fertility, and the purity of Kant’s argument from the universal law test to the content of the Moral Law. For each strategy, I will grant various parts of Kant’s argument for the sake of argument, to focus his attack upon another.

The First Set of Objections: Kant’s Theory of Maxims, Input into the Universal Law Test

As mentioned, Kant thinks that it is a part of the logic of action that an action is “done from” a maxim, a thoroughly general principle. It is clear that Kant believes each action is done

---

252 Korsgaard says the same thing (221).
from a single (properly specified) maxim. Call this the ‘singularity thesis’. In addition to speaking as though he accepts the singularity thesis, Kant is committed to it for a theoretical reason. The test of the Categorical Imperative is a test for the maxim of an action. If one and the same action has two (or more) properly specified, logically distinct maxims, the possibility arises that the universal law test will issue inconsistent practical judgments. It might happen that one (properly specified) maxim passes the test, and so is rationally permissible, and the other (properly specified) maxim fails it, and so is rationally forbidden. But this would mean that one and the same action is, according to Kant’s theory, both rationally permissible and rationally impermissible. To say the least, this is an implication that Kant would not invite; his theory is built to stave off such inconsistency. Following the Categorical Imperative is supposed to lead

---

253 How would Kant classify intentional or voluntary behavior that is done from more than one motive—an "over-determined" action? For instance, a person might volunteer his time both to help others and to help his reputation. Kant might think that one and the same behavior is to be described as two voluntary actions, each with its own maxim. But Kant might instead consider the maxim of an over-determined action as a complex principle that mentions both motives and the way in which the agent thinks of their relative importance: ‘Volunteer your time when it is possible to benefit yourself and others’. If Kant takes the second route, it seems that maxims can get very complex. I would suspect that Kant would see the instance of behavior I mentioned—volunteering time—as governed by two maxims. Then he can say that we ought to evaluate the behavior twice over: once to see if volunteering for other-regarding reasons is permissible, once to see if volunteering for self-regarding reasons is. After all, it is common enough for one and the same type of behavior to be permissible if done from one motive, but not another. See T.C. Williams (The Concept of the Categorical Imperative, 16-17). For the sake of clarity, I am going to think, for the time being, of cases of actions done from a single motive. We might call these ‘simple actions’. The singularity thesis ought to read ‘a simple action is done from one and only one properly specified maxim’.

254 To be careful, the objection might be restated in a slightly weaker way. Since Kant does not fully explicate the concept of a maxim, he has not provided clear necessary and sufficient conditions for (genuine) maximhood. A defender of Kant might respond to any instance of the above objection by claiming that it is not clear whether both of the so-called "properly specified” maxims really are properly specified, perhaps only one of the two candidates meets the actual necessary and sufficient conditions. But this response attempts to defend Kant’s theory by relying upon one of its weaknesses: its vagueness.
the agent to a unified vantage point—a practical point of view—from which to make practical decisions.\textsuperscript{255}

Two questions ought to be separated. The first is a theoretical question: Is the singularity thesis true? Is each action done from one and only one maxim? The second is an epistemic question: Is it possible, for non-omniscient creatures such as us, to discover which (single, properly specified) maxim an agent is truly (considering) acting from? As mentioned, Kant acknowledges that fallible human beings confront ineluctable obstacles to answering the epistemic question. We are going to set the epistemic question to the side. The first objection we are going to examine in this section challenges the truth of the singularity thesis.

An Objection to the Singularity Thesis

The “same behavior” can be subsumed under “more than one maxim.” As Sullivan puts it, “there is not a one-to-one correlation between maxims and behavior” (\textit{Immanuel Kant’s Moral Theory}, 160-162). John Hare expresses the objection in a common way, “there is in principle an infinite number of different maxims of different degrees . . . from which the same action can be said to follow” (\textit{Moral Gap}, 8). A person attempting to specify the maxim has a myriad of “candidate” maxims to select from. Which single one, from among them, ought she to choose?

The objection arises because it is difficult to see how Kant could answer, in a principled way, the question I posed earlier. The maxim of my action is to be derived from my precipitating conditions through a process of stripping away singular reference and generalization. The hard question for Kant is what level of generality to aim for. Once we begin to generalize away from

\textsuperscript{255} According to Kant, freedom is, as it is sometimes said, “to will only one thing.” And so, the unity of practical reason is necessary for freedom. The singularity thesis has to be true for practical reason to be unified and so for practical reason to predictably issue consistent practical judgments.
my particular motive, for instance, at what level of generality do we stop?256 We needn’t take the
guise of a trigger-happy, sharp-shooting skeptic to wonder what Kant would or could say; a well-
intentioned disciple of Kant is likely to want a bit of practical guidance at this point. Identify for
a moment with the would-be disciple of Kant and try to figure out the maxim of my giving Effie a
passing grade. How should my behavior be described? Say I give Effie a ‘D’, I can be
described—accurately—in any of the following three ways (ordered from more specific to less):
as ‘giving her a D while sitting in my office’, ‘giving her a D’; ‘giving her a passing grade’, and
‘giving her one letter grade higher than her academic work deserves’. Which of these accurate
action-descriptions is the one to be included, in imperative form, in the (genuine) maxim of my
action?257

Motives and circumstances often admit of the same flexibility. And Kant seems to agree.
In his examples, Kant himself often describes an agent’s motive in one instance, generally, as
“prudence” in the next, more specifically, as “a desire for money.” The agent is acting from a
desire for his own benefit, and he is acting from a desire for money. He simply takes having
money as in his own benefit. Which description, the more general or the more specific, belongs
in the properly specified maxim? Kant needs some principled way of stopping the process of
generalization, for each element of a maxim.

If Kant doesn’t give us a principled way to derive maxims, he opens himself up to the
accusation of cherry picking: choosing from among the myriad of possible specifications, the

256 And notice, even if you happen to act from a thoroughly general principle, one which makes no singular
reference, there is no guarantee that each of its elements is specified at the level of generality Kant’s theory
requires. We can’t tell until Kant tells us, and correctly, what level of generality we ought to aim for.

257 Perhaps I have always framed the issue, to myself, as a question about whether or not to fail Effie. But
even if that is true, does that determine that my maxim ought to describe my action more specifically as
‘Don’t fail’? Or more generally, as ‘Increase the grade’?
“maxim” that is most likely to get him to the results he antecedently wants. But whatever benefits Kant receives from giving himself this freedom to cherry pick, he pays for in different ways. Since Kant doesn’t give us a principled way to generalize away from the manifold details of the case, a critic might say that, for all we know, maxims could conceivably turn out to be exceedingly specific or exceedingly general. If exceedingly specific, the threat is that they would be so long and complicated to be of no use in practical deliberation. Also, they could include irrelevant details. (If Santa Claus tells a lie, is it “gray-bearded fibbing”? If maxims are exceedingly general, on the other hand, they could turn out so general that it is hard to make the case that the agent is rationally committed to that maxim: “I don’t advocate lying generally, only lying to save people from embarrassing truths.”)

To overcome this generality problem, Kant might appeal to his analogy between the process of specifying a scientific law and specifying a maxim. If a maxim includes the psychological states that actually play a role, at whatever level of the agent’s consciousness, in leading him to action, then it is possible that Kant could limit the influx of properties into the description of the circumstance-type, and so into the maxim, and in a seemingly principled way. A maxim won’t be too big. (If this number of properties wasn’t too big to play a role in leading you to action, it must not be too big to include in a principle of action.) And it won’t include

258 Herman calls this a socialization process, by which we come to learn a “moral range” of salient features. We subsequently bring this experience and knowledge to new situations. As I see it, an appeal to moral education is appropriate if we are talking with someone who is rationally committed to moral norms but is being obstinate in the present circumstances: ‘C’mon, you know better; you can’t tell a lie to hide another lie you told’. But if what we are doing is arguing that any rational being is rationally committed to moral norms, then an appeal to a process of moral education is of no use, argumentatively.

259 Schneewind considers a version of this objection—a critic wonders whether a lie told by Schneewind is an instance of ‘gray-bearded lying’—and calls the critic “perverse” (“Autonomy, Obligation, and Virtue”). But it is Kant’s failure to indicate how to stop what are obviously irrelevant details from spilling into properly-specified maxims that generates the objection. The critic needn’t deny that it is obvious they don’t belong. That’s what makes the objection a reductio ad absurdum.
what we intuitively see to be irrelevant details. Those aspects of an agent’s circumstances that are not playing a role in leading the agent to consider behaving are not to be included. In the process of scientific law-specification, we attempt to whittle down to the properties of the circumstance that are causally relevant, so too we might attempt to whittle down to the elements of the circumstances, or the agent’s perception of them, that are causally relevant to the agent’s behavior.

This response to the stark objection amounts to keeping a “tight” connection between the maxim of an action and the agent’s precipitating conditions. This same response might help a Kantian’s confidence in the singularity thesis. If there is a single true explanation in terms of the agent’s own psychological states about what is inclining him to act, then there is a single maxim. But if Kant takes this route, I think his tendency to favor greater generality is called into question. Not only that, it calls into question his claim that a maxim lacks singular reference. There are occasions, presumably, in which I conceive of my behavior in terms of singular reference, and irreducibly so. So, how can I be said, in these cases, to be impelled to action by a thoroughly general principle? The claim that maxims are thoroughly general principles is in this way contrary to the thrust of his analogy, which tells us to stop generalizing at the point we come to the casually relevant factors. What if a motive with an ineluctably particular object brought me all the way to action? Or what if the principle I am genuinely reasoning from includes singular reference?

Kant says that an intentional or voluntary action is ‘done from’ a maxim. The phrase ‘done from’ sound intentional-causal, and the account I just described treats it in this way. So we might say that an intentional-causal reading tends to disfavor the claim that maxims, if taken as impetuses to action, are always thoroughly general. Kant might change his tactic and argue,
instead, that a maxim is always “attributable” to an agent who has acted from a particular reason. Perhaps, he might say that acting from a particular desire to benefit Paul—or from an intention or principle that makes reference to him—*rationally commits* Paul’s father to a general principle that lacks singular reference, and so does not mention Paul. He might say that we are able to attribute the maxim to Paul’s father without it having made an entrance into his practical deliberation. (Kant, of all philosophers, is willing to read into an agent’s practical activity a range of volitional commitments.)

But that opens up the level of generality problem again. And, in any case, it’s not clear why we should think that voluntary action carries with it this implication. John Hare says, “it is not true that all practical reasons are derived from universal principles or laws, if this is taken to mean that individual or singular reference has to be excluded. . . we can agree that giving practical reasons requires some degree of generality, for otherwise we would not be giving reasons” (372). Hare usefully distinguishes between different “term positions” within a principle. In the example, imagine that Paul’s father has committed himself to the principle ‘Help Paul with his homework in the evening, after work’. This principle is general with respect to the

---

260 The analogy might be to the logical consequences of our beliefs. At least in some cases, we attribute the logical consequences of a person’s actual beliefs to him even if he does not have the further belief “in his head.” For instance, I might naturally attribute to George, after he has expressed the belief ‘boys are always smarter at math than girls’, the belief that he is smarter at math than every girl in the world, and then ask whether that is consistent with the evidence he is clearly aware of, the existence of female math professors, for instance. We often speak as though a person is guilty of an inconsistency between a generalization he expresses and the examples he hadn’t considered. It seems there are limits to this practice. Not just any logical consequence of my beliefs is attributable to me. But some are.

261 Some philosophers have objected that universalization, at least in some cases, forces the agent to engage in “one thought too many.” The standard example is of a loving husband who opts to save his drowning wife instead of a drowning stranger. He should not, the objection states, have to appeal to put his desire to save her to a test. That one person is his wife, and the other is not, is reason enough, without the need to universalize his maxim, to justify his behavior. Kant’s theory, it is said, ignores the special place in our practical deliberation of deep personal connections or special obligations.
action-type (‘help with homework’) and circumstance-type (‘in the evening, after work’). But the object of benefit, Paul, is particular. Why suppose that the father is committed to a more general principle that refers to sons generally? Or that he is not acting *voluntarily* unless he is? It’s not as though we would have a hard time understanding Paul’s father as an agent if he did not. It seems Kant is guilty of a fallacious slide from reasons to weak universalism.

The Upshot

Kant has to choose whether to favor more general or more specific descriptions. But troubling questions arise however Kant chooses to answer this challenge. If we take our cue from Kant’s discussion of his own examples, we ought to favor more general specifications. But if we favor more general descriptions in the maxim, it is often harder to argue that the agent is led all the way to action by that highly general principle. The particulars of an agent’s situation, or a highly specific motive, might have played an important role in bringing him to act. It might be my distress that leads me off the path of duty, and not a consistent and generalized egoism. In such cases, Kant would have to argue that acting for highly particular reasons implicates the agent in willing a highly general principle, the maxim. That will take some arguing, otherwise it will seem that he is attributing principles of choice to agents who are not committed to them.

So, we might then encourage Kant to favor less general maxims, so that he doesn’t have to argue for this type of attribution. But, there are also concerns if Kant goes the other direction. In many cases, the more specific the maxim, the harder it is to argue that the maxim cannot be willed, consistently, as a universal law. In fact, if we follow Pogge’s analysis of a maxim, and we

---

You might ask for a loan from your friend because he is the only person who will give it to you. Why, then, think that the maxim of the action is the highly general principle ‘Make a lying promise to get what you want’?
include elements of the agent’s circumstances in our specification of a maxim, these (even if described very generally) will often play a role in making a maxim pass through the universal law test. If the maxim of a particular instance of a lying promise were, ‘Make a lying promise in the event that you are in financial straights and want out’, then the subsequent universal law would kick in and compel obedience in the imaginary world only in those cases in which people are in financial straits. But this is a limited number of cases. Many promises have nothing to do with finances. People have been known to promise to be on time to meetings, or to bring cookies to a picnic. Presumably, the general institution of promise-making could survive its local abuse by people who have hit hard times financially.

In summary, the objection to the singularity thesis is that there is no principled way to determine the appropriate level of specificity or generality in the (genuine) maxim. This objection is reinforced by Kant’s own tendency to describe “one and the same” maxim in more and less general terms. And the upshot for Kant’s general theory is that multiplicity in maxims makes for, or at least threatens, inconsistent (moral and rational) prescriptions. Inconsistent prescriptions leaves the agent without a consistent answer—and so, without any real answer at all—to the questions that the Categorical Imperative is supposed to answer for an agent: How am I permitted, obligated, or forbidden to act?

263 Sullivan remarks that it is not clear that these problems are specific to Kant. But this problem is specific to any theory that makes the practical justification of an action dependent upon the formal properties of the principles from which the action is done. Such a theory makes it of central importance that an action is done from a specifiable principle. A theory that ties the practical justification of an action to its ability to bring about the satisfaction of an agent’s desired ends does not have this problem. The Humean theory is such a theory.
The Second Set of Objections: The Legitimacy of the Universal Law Test

The second set of objections has to do with strong universalism, the claim that an agent is rationally committed, for every action, to willing the maxim of the action as a universal law.

What’s the case against strong universalism? Largely, it’s that the claim is extravagant. Kant speaks of the agent as “willing” the maxim of his action as a universal law of nature. If we translate ‘willing’ in the standard Kantian way, as a matter of setting oneself to pursue a material end, then Kant is claiming that in acting intentionally or voluntarily an agent is setting himself to pursue the creation of a world in which his maxim is a universal law of nature: each action has, we might say, an “eschatological direction.” Your behavior reveals what kind of world you are willing to create. To act for any reason whatsoever is to reveal an aim, or at least a willingness or openness, to create a world in which everyone does as you are doing.

This gives sense to the Kantian talk of moral behavior as “creating a kingdom of ends”: Kant’s conception of a good, well-functioning society. When you treat a person as an end in herself, you reveal that you are pursuing the creation of a world, a kingdom or community, in which this is always done. When you abide by the (just) rules of your society, you are aiming to bring about a harmonious social world. Human behavior can be critiqued from the perpective of “kingdom-building.” Is what you are doing creative of a good, well-functioning society? Is it, in the very least, consistent with the creation of such a world?

Strong universalism, I would think, is a claim about what is attributable to an agent, not what he has before his mind when he acts. I would find it hard to believe that people are always governed by this type of “eschatological vision” each time they act. Perhaps good people do
commonly think in these far-reaching, social, “kingdom-building” terms, or are at least sensitive to the social implications of their behavior. But that’s another matter; we’re talking here about whether such a vision is a part of the logic of action. So, it seems that Kant is going to need to argue that it is legitimate to attribute to you.

Critics have commonly claimed that Kant is guilty of a fallacious slide from the existence of practical reasons to strong universalism. Alternatively, they could attempt to argue that Kant is simply building moral content into strong universalism and so is guilty of a question-begging move. In response, Kant would likely appeal to two, connected considerations. First, Kant’s justification for strong universalism is that a right understanding of the nature of the will implies it: the will is designed “by nature” to legislate universal laws; it is a “causality,” and by nature “a causality operates according to universal laws.” This seems to reflect another aspect of Kant’s commitment to the natural law tradition. Various aspects of the world, in this case, of our psychology, are said to have an in-built purpose that serves to indicate how it ought to be used. Kant rejects some aspects of the natural law tradition, those which claim that an in-built natural teleology sets for us substantive goals that we ought to pursue, because he thinks these claims are inconsistent with human freedom. But Kant thinks that we are, as it were, built for the exercise of freedom. (Korsgaard, as I see it, attempts to stay away from natural law ideas.)

Second, Kant also makes an analogy between practical and descriptive laws. So, perhaps he thinks that the properties of scientific laws are also the properties of practical laws. Kant has a Newtonian conception of scientific laws; what makes a scientific law “law-like” is its

---

264 Korsgaard saves the Kantian idea by ratcheting it down: a pair of friends, in treating each other thoughtfully, create a “republic of two.”

265 See Harman (Nature of Morality, 76), and Cullity and Gaut (Ethics, 21).
universality, it makes no singular reference and it hold in every case. So, to properly call a practical rule a “law” it must also be fully universal.

But why think that a principle of choice—an action-guiding principle—must meet the criteria that a scientific law—a descriptive law—must meet? As Nagel remarks,

So far as I can see, choosing freely in a law-like pattern is merely a way of mimicking causality; if I always put on my left sock before my right, that does nothing to establish the causality of my will, so why does the categorical imperative do any better? There has to be something more compelling about the demand for universality than this (“Commentary,” 202).

I agree.

The Upshot

Notice, the truth of both weak and strong universalism are necessary for Kant’s argument to be sound. An agent can be accused of an inconsistency between two volitional states only if these volitional states can properly be attributed to him. The objection is, at least in a wide range of cases, these attributions can’t plausibly be made. Kant seems to be guilty of an analogue of the straw man fallacy: he attributes to agents volitional states that they do not have, and then calls them irrational for having them.

If these questions are left unanswered, it will appear that Kant’s general argument and theory is under-motivated and so has a gaping hole in it. If weak and strong universalism are false, then Kant is left with an “empty self”—or at least an “emptier” self than he is aiming for. According to Kant, a human self consists of the capacity for reflection, along with the rules of rationality, theoretical and practical. These rules provide the agent with a vantage point from

266 For an answer, see Henry Allison (Kant’s Theory of Freedom, part III). See also Thomas E. Hill, Jr., “Kant’s Argument for the Rationality of Moral Conduct”).
which to make decisions. But if weak and strong universalism are false, then they are not among the rules of practical reason, and what is left of the picture Kant has been drawing of the self is only the capacity for reflection and a principle of instrumental reasoning. These do not promise to be enough to compel moral behavior.

To Kant’s credit, many philosophers have found the universal law test to be a legitimate test of normative ethics, a principle—or perhaps the primary principle—of justice. But thinking that the question ‘What if everyone acted like this?’ is an important part of moral deliberation is consistent with disagreeing with Kant’s belief that the thought experiment is a test of the sufficiency of a practical reason itself.

The Third Set of Objections: The “Fertility” of the Categorical Imperative

Let’s grant the singularity thesis and weak and strong universalism, for the sake of argument. If we do, does the universal law test successfully forbid each immoral maxim? The universal law test might be too weak to rule out each of them. I will raise an objection that claims that it is.

There is reason to think that the universal law test is too strong—“overly fertile.” It seems to forbid maxims that are neither immoral nor irrational. It forebears, for example, the maxim ‘For the sake of your financial benefit, pay off your credit cards before they incur a single penny of interest’. This maxim generates a practical inconsistency. A world in which everyone successfully obeyed this maxim would be the death of the credit card industry, and in that event

---

267 Hegel remarks that the principle ‘Succor the poor’ generates a contradiction in the conception, though it is clearly not immoral. In a world in which everyone succored the poor, there would be—Hegel presumes—no poor, and so one would not be able to will the maxim as a means—to help people—in such a world. The beneficent maxim, if it were a universal law of nature, would render the maxim useless as a means; and so to will it—Kant would be forced, implausibly, to say—is to be guilty of a practical contradiction.
you could no longer successfully will the maxim as a means to your financial benefit. But fiscal prudence isn’t irrational or morally wrong. The implication is that the Categorical Imperative is too fertile: it rules out behavior that, intuitively, it should not.

In this section, I am going to discuss several versions of the opposite charge: the universal law test of the Categorical Imperative is too weak: there are immoral maxims it fails to forbid. The diagnosis is that the universal law test is insufficiently fertile because of its being a formal principle. Within the literature, fertility objections come in many forms. In its strongest form, the objection is that the Formula is “utterly empty”: it rules nothing out and is not even “minimally demanding.” In a few of its more moderate manifestations, the objection claims that the Formula of the Universal Law does not secure some aspect or other of the “matter” of the Moral Law, which includes both the second and third formulations of the Categorical Imperative, on the one hand, and a number of principles he calls “categorical imperatives,” on the other. Among the little categorical imperatives are principles that proscribe lying, committing suicide, ignoring the plight of the poor, and so on. The moderate versions of the empty formalism objection claim that at least some of these particular derivations don’t work.

I want to make it clear that the first two objections—against the singularity thesis and against Kant’s legislative universalism—are not always distinguished from what I am calling “empty formalism” objections. Sometimes the reason that the Categorical Imperative is regarded as empty, or insufficiently fertile, is that the critic denies the singularity thesis and runs several possible specifications of a single action through the universal law test until he finds a result that doesn’t fit with Kant’s (or his own) moral judgments. And sometimes the objection derives largely from the fact that the commentator has denied weak or strong universalism and so has gutted the universal law test. (Allan Gibbard seems to do this (“Morality as Consistency”)). It’s
not hard to pass maxims through the test if the heart of the test (weak and/or strong universalism) are denied. In this section, I am going to consider objections that do not deny the singularity thesis or weak and strong universalism.

To foreshadow some of my discussion in the next chapter, in *The Sources of Normativity*, Korsgaard herself refers to a version of the empty formalism objection that has “a kernel of truth,” but she doesn’t describe it. I think it is fair to say that Korsgaard’s arguments are centrally constructed to overcome each, and she is especially concerned to vindicate the Formula of the End in Itself and the Formula of the Kingdom of Ends, whereas she does not emphasize the other moderate empty formalism objections.

The Source of Suspicion

Empty formalism objections arise because Kant thinks both that the Categorical Imperative is the Moral Law and that the Categorical Imperative a (thoroughly) formal principle, the second claim Kant centrally makes about the Categorical Imperative. When Kant speaks of the Moral Law, he is thinking of particular, substantive requirements. Kant thinks that the test has particular implications: it implies what we regard, or at least what he regards, as immoral to be rationally forbidden. This principle is supposed to have independently specified moral implications. Kant also thinks that the Categorical Imperative itself is a (purely) formal principle—the practical principle of non-contradiction. A human being is, in essence, a rational creature, and so, it is only the principles of practical rationality itself that can compel commitment in an autonomous (and not a heteronomous) way. None of the principles of practical reason itself contain any substance, except whatever criteria are implied by the concept of a law. The worry is that these two parts don’t fit together.
Let me illustrate. As we saw earlier, Kant himself makes an analogy between the (theoretical) principle of non-contradiction (and its corollary for belief formation) and the Categorical Imperative. The theoretical principle of non-contradiction suggests a principle for belief formation: ‘Do not believe both P and not-P (or anything that implies both P and not-P)’. (I’ll set aside the ‘or anything that . . ’ phrase, for the sake of clarity.) If you believe both P and not-P, then you know (according to the theoretical principle of non-contradiction itself) that one of your two beliefs is false. And this suggest that (and here is the corollary for belief formation) you ought to head in the direction of a consistent belief set, which means giving up at least one of your inconsistent beliefs, if not both.268

The logical law of non-contradiction plays, so to speak, a negative role in belief formation: it tells you, of some particular sets of beliefs—those that are inconsistent—that they are rationally impermissible. By itself, it does not tell you, positively, what to believe. For instance, this principle for belief-formation does not tell you, if you believe both P and not-P, which of the two beliefs to give up; it simply tells you, other things being equal, to give up at least one of them. Perhaps if other epistemic considerations are brought in, it will become clear that it is more rational to give up one belief rather than the other; but that consideration would be another consideration, logically independent of the logical law of non-contradiction itself. Consistency in belief does not guarantee truth in belief. And, since there is more than one set of consistent propositions, even if the principle of non-contradiction can help us to weed out many

268 What about paradoxes? That is, what about cases in which two, inconsistent claims each seem to have sufficiently compelling support? Perhaps in some of these cases, it is warranted to believe inconsistent claims. Could Kant say so? It doesn’t seem so. What about the practical analogue: cases in which two, inconsistent principles both seem to call for affirmation, even though they contradict? Perhaps Agamemnon at Aulis is in such a position; he has absolute practical commitments to both his army and his daughter, which lead to conflicting and “deep” obligations. Must he give up “unconditional” support for one of them? Kant seems to demand that he does.
possible belief sets, the ones that are not consistent, it can’t, by itself, pare the possible belief sets down to one. We’ve got to get guidance from somewhere else: epistemic principles, for instance.

Likewise, say that you happen to be willing (setting yourself to pursue) both W and not-W. The practical principle of non-contradiction tells you not to will both W and not-W (or anything that implies W and not-W). (Again, set aside the ‘or anything . . . ’ clause.) In the event that you find in yourself such a volitional contradiction, this law does not tell you which volitional state to give up, W or not-W; it simply tells you to give up at least one of them. Other considerations might make it more rational to give up one willing rather than the other—perhaps one of them coheres better with your most cherished ends, or better tracks the good; but these are considerations that are logically independent of the practical principle of non-contradiction itself. And, since there is more than one consistent set of volitional states, the practical principle of non-contradiction can’t help us to pare down the possibilities to but one.

Among the sets of consistent volitional states are, it seems, immoral sets. In Alan Gibbard’s words, “there are many possible ways of being consistent. Most of them would be crazy, and some are downright vicious” (144). If Caligula were to have the sole life plan of maximizing the sufferings of others, he would be pernicious partly because of his single-minded consistency.

Take an example. Imagine a person who has contradictory volitions—he wills (to guide his conduct by) the principle that all citizens ought pay their taxes, and he wills in a particular case (selfishly) that he not pay them. (This example is an instance of what Kant calls making an exception of yourself.) If confronted with the inconsistency, the citizen could render his volitional set consistent, and so appease the practical law of non-contradiction, by giving up either his general (citizenly) volition or his particular (selfish) volition. But one of the ways he
might become consistent is by choosing the immoral volition and rejecting the moral one. If the Categorical Imperative is a formal principle, then how—it is natural to wonder—is it to force this citizen (or anyone else) to veer to the (morally) right instead of to the (morally) wrong? The citizen has the option, as it were, of veering right—to the volitions with "genuine moral worth"—or of veering left—to the willings that are forbidden by the Moral Law. Why must he “veer to the right”?

It looks as though we are going to have to hope that the person’s citizenly principle has a grounding in his heart’s desire, or something along those lines. Or we’re going to have to seek out realist standards that require that his willings track the (morally) good, or something along those lines. In short, we need to be able to appeal to something substantive. Consistency by itself won’t favor one principle over the other. A requirement to have self-consistent principles, along with a requirement to have mutually consistent principles, is not robust enough. If the Formula of the Universal Law is equivalent to, and so nothing more than, the practical principle of non-contradiction, Kant’s theory fails.

None of this is to deny that consistency has real bite. The purpose—or, at least, a central purpose—of volitional activity, presumably, is to affect the course of things, often to bring about states of affairs that we care about. (Kant does not deny this. Recall a necessary condition for rational behavior is that it be aimed at an end, and Kant does think that our material ends are things we desire.) If willing is a matter of setting oneself to bring about a state of affairs, then we

---

269 See Frankena (Ethics, 25).

270 See Gibbard’s discussion of “substantivalism” (“Morality as Consistency”).

271 Velleman appears to deny this. He claims that the “constitutive aim of action” is to make sense to ourselves. It seems to me that “the constitutive aim” of action, if it has one, is to affect the course of things, in ways that we care about. Only one of the things that we care about is making sense to ourselves. Many ends that we pursue are states of affairs that are not self-referential (“What Happens”).
have—if you permit me to speak in a way that sounds question-begging—a practical reason to set ourselves to mutually consistent ends, and so to principles for action that are generally consistent with each other. Gibbard gives the example of an author with the following three thoughts, each of which is, by hypothesis, true of the author: I intend to write, I intend to go to the beach, I can’t write at the beach. The author suffers from a volitional inconsistency, and one he ought not shrug off. If the author is to write, he must forego his intention to go to the beach. If he goes to the beach, he must forego his intention to write. At present, he has a set of intentions at least one of which is going to be frustrated. He is, as it were, divided against himself. And so, he must re-think his set of intentions; if he aims to act upon both, one of the intentions will be frustrated. He is then working at cross-purposes.

To get back to the empty formalism objection. As I understand it, it states that if a maxim is simply a principle of action—or a plan of action, or perhaps even a life plan (as some defenders of Kant, such as Sullivan, prefer)—then it seems as though there are malicious principles—or plans of action, or life plans—that are fully consistent, but immoral. Consistency requirements aren’t sufficient to compel moral behavior.

Kant’s Response

Kant might argue that the empty formalism objection attributes to him a different (and perhaps weaker) principle of consistency than that exemplified in the Categorical Imperative, and

---

272 Is a principle of volitional consistency absolute? I wouldn’t think so. It strikes me that there are cases in which it is practically rational to adopt principles of choice, say, that are inconsistent. Perhaps this is what got Agamemnon in trouble: he make an absolute commitment, as a father, to his daughter, and an absolute commitment, as a general, to his fatherland. Is the mere possibility that these could come into conflict a sufficient reason for him to modify at least one of his practical commitments? I don’t think it is.

273 Gibbard has another good example of a binge drinker who resolves to change his ways, but goes to a party in the full knowledge that he can’t resist drink once there (“Morality as Consistency,” 145).
then the objection claims, lamely, that that different (and weaker) principle does not imply the Universal Authority Thesis.

In Kant’s defense, the Formula of the Universal Law does not state ‘Do not will (set yourself to pursue) both W and not-W at one and the same time’. It states that one must will a maxim only if it can simultaneously be willed as a universal law. Kant does not advocate merely a “bare” principle of non-contradiction, so to speak, but a principle of non-contradiction that is sensitive to the substance of the concept of a law. Kant includes in the Universal Law Formula what he takes to be several aspects of the concept of a law. A natural law makes no singular reference and it holds in every case.

This requires, I suppose, Kant to take a bit of the blame for the misunderstanding. (Though let’s not dwell on it, except to clarify the argument.) He has to say that the Formula of the Universal Law is a principle that forbids inconsistency in willing, but it is not itself, as he has intimated, the more general practical principle of non-contradiction. The generic principle of non-contradiction forbids any willing of W and not-W. This includes the types of contradiction that the Categorical Imperative seeks out, but also includes garden variety inconsistency within my personal will: my willing, for instance, both that I play soccer every Sunday and that I refrain from physical exercise on the weekends.

A Kantian might hold out hope that the thought experiment in the universal law test Kant has sufficient content, grounded in the concept of a law, to overcome the empty formalism objection. This defense suggests that Kant’s talk of the Categorical Imperative as the practical principle of non-contradiction does not tell the whole story. It is necessary to include the concept of a law (or something else). But this simply takes us back to the challenges raised to Kant’s legislative conception of practical reason discussed in the last section. Protecting Kant from this
version of the empty formalism objection requires committing to other highly controversial elements of Kant’s theory, namely, weak and strong universalism.\textsuperscript{274}

The Fourth Objection: Impurity in Kant’s Arguments

This argument, though, seems to violate one of Kant’s own, self-imposed strictures. Kant believes in the purity of the Moral Law; his argument for the universal authority thesis is not to be “infected” by appeals to practical anthropology: contingent truths about biologically human beings. Given this constraint, the volitional states that generate a contradiction in the will must be valid for each and every practicing agent and so grounded in the nature or conditions or activity of rational agency. But Kant seems to let contingent truths slip in.

No doubt, we are inescapably needful. I can’t see any route to god-like self-sufficiency. But even so, our needfulness is a contingent fact about us. Also, willing the bad Samaritan maxim, no doubt, violates the golden rule, is very unfair, and so is morally objectionable. (Who doesn’t chastise his children for this?) But Kant’s primary claim within his argument is that so willing is rationally vicious: a violation of the rules of rational willing, which forbid contradictions among our volitional states.\textsuperscript{275} (‘Sometimes give help’ is an imperfect duty for each of us.)

Is Kant guilty of slipping from the technical meaning of ‘human being’, according to which any rational being is a human being, to a more common meaning? Kant could say that every human being is rationally compelled to adopt the maxim ‘Help others’ (an imperfect duty).

\textsuperscript{274} For a defense of Kant against the empty formalism objection, see O’Neill (“Consistency in Action”).

\textsuperscript{275} On this interpretation, Kant’s argument seems to me to turn on a slippery move: the maxim subtly changes from ‘Deny financial assistance’ to ‘Deny any type of assistance’. As mentioned earlier, Kant is insufficiently clear about how general a properly specified maxim ought to be.
But he would have to be using the term ‘human being’ in its common and not its technical sense. In the technical sense, God is a human being, and He is self-sufficient.

There are serious concerns even if we grant these controversial elements of Kant’s theory. For instance, contrary to Kant’s argument, it seems that it is possible for a lying promise to be instrumentally effective in a world in which the maxim of the action is a universal law of nature, even on Kant’s own construal of the maxim. If we include in the maxim the genuine precipitating conditions, these will often play a role in limiting the instances in which the maxim compels action. If one of the genuine precipitating conditions inclining an agent to consider a particular lying promise is the motive to get out of financial straits, then the corresponding universal law will only lead to lying promises by the down on their luck. The global institution of promise-making and promise-keeping is able to survive its local abuse by the down on their luck. Subsequently, the universal law test is not able to catch the lying promise-maker in a contradiction in the conception. This is to say, the answer to the question how maxims are to be properly specified has implications for the plausibility of Kant’s legislative conception of practical reason (if they are more general, they are hard to attribute to agents) and for the empty formalism objection (if they are more specific, they will tend to pass through the universal law test more easily.)

In his examples of volitional contradictions, Kant often appeals to premises that do not seem to belong, given the strictures Kant places on his own argument. Kant aims for the Categorical Imperative to be a test of rational agency, as such. Any being, no matter what desired ends he pursues, is subject to the demands of the Categorical Imperative, in each of its three formulations. The way Kant intends to reveal this is that the universal law test implicit within the first formulation is supposed to imply, by some route or another, the second and third
formulations, as well as a wide range of “common sense” moral obligations. Subsequently, in these derivations, Kant should appeal only to premises that are (if not tautologically derived from the conception of a law) a part of “rational anthropology” and not to premises that describe what are, at best, contingent truths about human beings and the world we happen to occupy. The objection is that he often does appeal—improperly—to mere contingent truths about human beings and our world.

Take an example. Recall that Kant wants to get to the conclusion that a lying promise is forbidden by the universal law test. But, in a logical sense, it is possible to conceive of a world in which the institution of promise-making survives despite a law of nature that compels people to make lying promises any time the relevant precipitating conditions obtain, or even anytime doing so is merely expedient. If we consider an agent to be a creature with the capacity for reflection and a set of desired ends, but then let our imaginations go free, we could imagine plenty of worlds in which the maxims Kant condemns pass the test of the universal law. Imagine a world with plentiful goods, in which the relevant precipitating conditions (dire financial straits) rarely, but occasionally, obtain. In such a world, which is logically possible, people to whom lying promises are made—because they happen so infrequently—have too little experience of them to become strongly suspicious of the institution of promise-making or to think such a promise is nothing more than a “vain pretense”—even if it always is one. Also, the institution of promise-making might survive—despite universal lying promises—in a world in which the majority of people suffer from very poor memory and so consistently forget having been duped each earlier time they came across lying promises. Or even in a world in which there is a wide spread tendency not to prejudge a person and to practice forgiveness “seven times seventy” times. Kant seems to be
including in his employment of the universal law test contingent, empirical facts about our common levels of poverty, memory, and personal impatience with each other.

Does Kant lack imagination to think up these possible, albeit strange, worlds in which things are different? I doubt it. What these examples suggest is that Kant is making assumptions about what the imaginary world and the various agents within it are like. Perhaps Kant intends for us to make our judgments of “conceivability” from a set of presuppositions about human mental ability. Or perhaps we are to make these judgments from a perspective that holds firm some features of our present world—scientific laws and other empirically verifiable truths. Or perhaps both. In short, the imaginary world in which the maxim is a universal law shares much in common with the world in which we live and act.

Some commentators have suggested this reading of Kant. For instance, Pogge says, in “The Categorical Imperative,”

Kant takes for granted a general understanding of the laws of (human) nature or of the permanent conditions of human life . . . he consistently invokes general facts: We have the need for the love and sympathy of others (423), we have sufficient memory to learn from experience (which we are disposed to do), and we can use language to deceive one another (403, 422)” (194).

So, what Kant is then thinking is that the institution of promise-making could not survive in a world in which making lying promises is a law of nature and which includes creatures much like us—with our level of memory, our general concerns, our level of patience, our emotional needs and status anxiety, and so on.

Pogge does not seem to think that the universal validity of the Categorical Imperative is compromised by including these contingent facts about us. But, contrary to Pogge, it is a bit puzzling how this can be. If contingent facts about human beings are included in the applications
of the test, the test of the Categorical Imperative seems to be infected with a kind of contingency that Kant claims to want to eradicate from his fundamental moral principle. It seems fair to say that our particular level of memory is a contingent trait. Having particular emotional needs is, too. By this route, contingency seems to seep into Kant’s argument for the practical authority of the moral law, and his is an argument that is constructed with the intention to keep it out.\(^{276}\)

I see two ways to attempt to defend Kant. One way is to argue that he appeals only to features of our humanity that properly count as conditions of agency. For instance, he might argue that a being cannot be an agent without the faculty of memory. To be a practicing agent implies, Kant might claim, aiming to bring about states of affairs presently but also in the future, and we are not able to do that without being able to remember the intentions and plans of action we have adopted. Likewise, it is an interesting question to ask whether a being without any needs whatsoever could possibly be an agent. There are certainly some interesting thought experiments to engage in here. (“Can we make sense of a being without . . . being at the same time an agent?)

But I don’t think this argument is going to accomplish what Kant wants. I suppose it is defensible to argue that if you cannot remember that you have adopted a principle of choice, then you’ll be hard pressed to govern your conduct according to it. To be a persisting agent, across time, with persistent intentions and practical commitments, you must have enough memory to recall these commitments. Also, arguably, some ability to learn from experience might be necessary for agency. But, that said, our particular level of memory, and the particular kinds of generalization from experience we make, aren’t conditions upon agency. And it is conceivable

\(^{276}\) I suppose the fact remains that the Categorical Imperative asks whether willing a maxim as a means and as a universal law is consistent. The problem isn’t that the imperative becomes a substantive rule that states ‘Do this’ or ‘Don’t do that’. It’s that Kant’s judgments about what a world in which the maxim is a universal law is necessarily like are not accurate. He’s ignoring possible worlds in which the maxim does function as a universal law, and in which the maxim could yet be willed as a means.
that the people in the imaginary world generally have human-like levels of memory, but are apt to forget specific kinds of events, such as getting cheated. (They are “cheated-forgetful.”) Or maybe their memory generally functions well, but they never want to hold a grudge and are very good at forgetting ills suffered. That is, it is possible that a general policy of forgiveness, if also commonly accepted, makes it possible for the misbehavior to persist. Kant’s theory is to apply to all rational creatures, as such; are very forgiving creatures not even minimally rational? And, moreover, the general strategy being examined here—we are claiming that Kant says that some level of memory is necessary for agency—wouldn’t work for other anthropological claims Kant appeals to. Status anxiety and the need for love, which Kant himself mentions, do not seem to be a part of our very rationality.

Sullivan suggests a second way to defend Kant. Take another example. As discussed, according to Kant, it is a violation of the Categorical Imperative to will a bad Samaritan act. This is because you are not self-sufficient, often in need of help. Therefore, you are rationally committed to willing that another person, someone whose help you need, not be a bad Samaritan to you. This rational commitment is due to a contingent feature of biological humanity, our lack of self-sufficiency. The capacity for rational agency—humanity in Kant’s technical sense—does not seem to entail a lack of self-sufficiency, that is, the need for help. (Would God be incapable of rational agency? A rational being might have aims without having needs.) Again, Kant seems to include contingent facts about us in the derivation.
Sullivan agrees with Pogge that Kant consistently included what we would generally describe as contingent claims in the universal law test. But the reason that Kant was not bothered by this is, Sullivan tells us, Kant is in some sense, a “natural law” moral philosopher . . . Kant believed that there is no way in which to apply the Categorical Imperative so as to generate many of our moral obligations, particularly positive duties to ourselves, without reliance on the natural law (Immanuel Kant’s Moral Theory, 150).

What we, with our contemporary attitudes toward scientific laws treat as mere contingent features of our world, Kant treated as, in some sense, special. The laws of nature, in this view, bear some type of necessity. Things are in nature the way they are for a purpose.

One interpretation of Kant fits with this idea. Kant’s system could be viewed as a rationalized form of Christianity according to which God, the king of the kingdom of ends, has ordained that human beings have various traits, such as a particular level of memory, a lack of self-sufficiency, the need for affection, and so on. These traits, though contingent in a logical sense, are intentionally built into the fabric of our being by our creator, God. So, though they are not logically necessary in the sense that any rational being has these traits—God does not suffer memory lapses or a lack of self-sufficiency, for instance—they are put in us for a purpose. In some sense of the word, they are not accidental features of our human nature. Human nature is chosen.

If this interpretation is correct, though, Kant has to give up an important element of his conclusion. The argument would not be able, even if sound, to establish the validity of the Moral Law for every rational being, as such. It would have to limit itself to aiming for the conclusion that rational beings with a lack of self-sufficiency are subject to the Moral Law.
I don’t intend to launch an argument against the philosophical and theological tradition that makes claims such as this. But if Kant’s arguments depend upon many and thick assumptions from the natural law tradition, or from the Christian tradition, Kant’s theory begins to take on new premises that need to be defended. Korsgaard, in fact, seems unwilling to advocate these aspects of Kant in her argument.

The reason that I am including this issue under the title of an objection to Kant’s theory is that many philosophers are not sympathetic with the natural law tradition. If Kant’s universal law test is dependent upon that tradition to generate practical judgments, it would loose its force with many contemporary philosophers who do not invest a special type of necessity in these claims, even if they are regarded as accurate descriptions of us. And, for our purposes, Korsgaard’s way of adjusting Kant’s own argument and theory manages to avoid these concerns.

Final Remarks

In the first chapter, I remarked that a Kantian might propose that he has what Frankfurt seeks, a proposal for how to account for the rational impotence thesis. According to Frankfurt’s subjectivist theory, the will centrally consists of motivational states, and it might seem difficult to understand how one set of motivational states could rule out other motivational states. At least initially, the will, construed in a subjectivist way, does not seem to have the right shape to categorically reject basic desires. My proposal suggests that norms of rational impotence are incorporated into the will by being elements of the objects of desire. This means that norms of rational impotence are not a necessary part of rational agency; an agent might not care about any personal ideals that include norms of rational impotence. If an agent cares in the proper way
about a personal ideal, it is able to occupy a stable and important place in his practical point of view—especially if it is supported by several of the agent’s ruling passions. A father’s concern for his son, and his resolution to be a decent father, might themselves be things he cares strongly about and is averse to changing in himself. They might also give him a sense of pride, and useful goals to work towards. His personal ideal, then, is also supported by a type of self-love and by a desire for meaningful pursuits. This, it seems to me, goes a significant way to explaining why the commands of his personal ideal, or at least some of these commands, confront him with such compelling force. This strikes me, as well, as a theoretically economical way of arguing for the intuitions surrounding the rational impotence thesis, and without incurring the manifold and tangled troubles of Kant’s system. Norms, with the relevant type of ‘must’—‘Do not treat this desire as the source of a practical reason’—are elements within a complex object of desire: the agent’s cherished personal ideals.

Korsgaard, though, is a significant dialectical partner for me. She agrees with Kant’s conclusion and with the general character of his argument. But she thinks it needs to be supplemented by an appeal to the concept of a practical identity. I regard a personal ideal as a normative standard that issues substantive practical advice, potentially about a wide variety of topics: how to behave, how to deliberate, how to feel, what character traits to have, and what (not) to treat as reason-giving. This seems to be what a practical identity is, too. Korsgaard’s only definition of ‘practical identity’ is difficult to parse, and she largely gets at the concept by way of examples and suggestive phrases, which suggest that a practical identity is “an object of aspiration” (238), which is connected to self-regarding emotions such as pride and shame, and is possibly a source of meaning for the agent. She thinks that the concept of a practical identity belongs in a Kantian theory, not a subjectivist theory.
To understand Korsgaard’s project it is necessary to recognize the theoretical constraints she places upon her own argument. Korsgaard accepts a litany of Kantian claims. Here are the most pertinent to my discussion. Korsgaard professes to be—with respect to practical reasons—a voluntarist, a constructivist, and a procedural (but not substantive) realist.\footnote{Korsgaard discusses voluntarism at (18-19, 21-27, 90, 104, 164-165); constructivism at (19, 35, 171, 246-247); and procedural and substantive realism at (35-37, 44-48, 112, 205-208, 245-246).} She thinks that practical reasons derive, ultimately, from principles of choice that are incorporated into the agent’s personal will by acts of self-legislation that are constrained by principles of volitional consistency. She also thinks, correspondingly, that a basic desire, as such, is rationally impotent: a motivational state, once recognized, presents “a problem”—“not a reason.” This reflects her commitment to the heteronomy thesis, the thesis that, among other things, basic desires are external to the will and so cannot properly “determine” what the agent does. All of this is very much Kant, and it firmly stakes Korsgaard’s claim to being, in a very deep sense, a Kantian about practical reason and the Moral Law.

Much of Korsgaard’s argument is driven by her commitment to Kant’s claim that the will, to be free, must be “entirely self-determining” (97). Korsgaard rejects theological voluntarism, realism, and subjectivism on the grounds that, respectively, the dictates of God, of a realist standard, and of desire are “external” to the will and so cannot properly be treated as ultimate sources of practical reasons. The considerations that theological voluntarists, realists, and subjectivists appeal to can become sources of practical reasons for an agent only by being endorsed by a proper act of will, an act of self-legislation, “influenced” but not “determined” by
any external cause. For ‘because God said so’ or ‘because I care about it’ to count, properly, as a reason to act, it is necessary for the will to endorse these considerations.

Following Kant, Korsgaard is deliberately anti-subjectivist. It is Korsgaard’s firm intention never to deviate from the heteronomy thesis, Kant’s claim that desires are external to the will and so are permitted to “influence” but not “determine” the operation of the will. She explicitly rejects even sophisticated forms of subjectivism, thinking that subjectivism fails, by its very nature, to answer all of the questions an agent needs an answer to; in treating basic desires as reason-giving, it encourages, she thinks, agents to end the reflecting self’s regress of ‘whys’ before the job of reflection is done: basic desires need justification, too. This, you might notice, is an intimation of the argument from full reflective agency, and it is driven by the heteronomy thesis: ‘because I want to’—even ‘because I care about it’—is never a proper regress-ending answer. According to Korsgaard, on the other hand, it is the will that “does the normative work,” independent of desire, of creating a valid practical vantage point, of “making laws unto itself.”

278 Korsgaard also makes the claim that an “ethics of autonomy” is “the only one consistent with the metaphysics of the modern world” (5). So, Korsgaard alludes to another argument, metaphysical in nature, that is also supposed to support her theory of practical reasons—or, as she puts it, of “normativity.” She says that we can no longer go back to the pre-modern worldview that value is “more real” than the physical world we experience. And Kant’s theory is the best theory that is consistent with the Modern Scientific Worldview and the “mystery of value” (1-5).

279 She calls this view the “reflective endorsement” method and attributes it to David Hume, J.S. Mill, and Bernard Williams. In her interpretation, Hume and Mill attempt, each in their own ways, to argue that the moral life is commonly an instrumentally rational way of life for a human being to live; living morally serves ruling passions that many human agents often happen to have. Korsgaard calls this the reflective endorsement method (Sources, chapter 3).

280 I challenged this move in chapter 3. I agree that a basic desire admits of a wide range of normative questions, and I think it is coherent to ask, of any basic desire, whether it is rationally potent. Even so, for all of that, the answer might be yes. The ability to ask a question “Is such and such an ultimate source of practical reasons?” is not a reason to think it isn’t. To think so fails to recognize that a theoretical account of what makes an ultimate source of practical reasons an ultimate source of practical reasons is not itself a part of the chain of justification.
Let’s see whether her theory is a more compelling, or at least more promising, proposal for handling the rational impotence thesis.
CHAPTER 6
KORSGAARD AND PRACTICAL IDENTITY

Kant’s argument for the universal authority thesis is complex. And the literature that concerns it, its defense and its criticism, is long and complicated. I have raised four sets of objections to Kant. Someone intent on defending Kant might hold out hope that a clear and distinct presentation of Kant’s argument and conceptual scheme, or at least a more thorough working out of it, would lead to a sound argument. Korsgaard doesn’t. She thinks that a sound argument for the validity of the Moral Law, and the will’s corresponding capacity and practical authority to disenfranchise immoral inclinations, requires important departures from Kant, “modifications” that she “has come to think are necessary” (91).

Korsgaard does aim for a sound defense of the universal authority thesis. One of her primary goals in Sources is to advocate for the voice of conscience and its “right” to tell us what (not) to do; she wants to provide an argument “to show that any reflective agent can be led to acknowledge that she has moral obligations” (125). Korsgaard agrees with the idea—central to Kant—that the logic of action rationally commits each practicing agent, as such, to the matter of the Moral Law, to be respectful and fair. Invoking both the words and the spirit of Kant, Korsgaard claims that “the value of humanity itself is implicit in every human choice” (122). To

---

281 Korsgaard requires of her interlocutor that he be “sincere and reasonable” and that he “does really want to know” whether the Moral Law is valid for him (16). I suppose these are the expectations we generally place upon interlocutors.
make a choice, to will anything whatsoever, is to be implicated, she thinks, in “valuing” yourself, other human beings, and the human community as a whole.  

One of Korsgaard’s primary modifications is an appeal to the concept of a practical identity. There is a question about what a practical identity is, about how much she means to include in its definition. It is clear from Korsgaard’s discussion that a practical identity is a normative standard that generates substantive practical advice about, among other things, how (not) to act and what (not) to treat as reason-giving; and so, a practical identity is able to function as a practical guide for behavior and volition. More to the point, recall that Korsgaard believes that there exists a proper “kind of test of normativity” for any motivational state: “as each impulse to action presents itself to us, we should subject it to a test of reflection, to see whether it really is a reason to act” (89, emphasis hers). A practical identity is such a test: “We answer that question by seeing whether the maxim of acting on [the impulse] can be willed as a law by a being with the identity in question” (113).  

I have questions about what else, if anything, Korsgaard includes in the definition of ‘practical identity’. She glosses it in the following way: your practical identity is “a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking” (101). This connects a practical identity to a sense of

---

282 Korsgaard has a tendency to construe her theory as a theory of “normativity” or of “value” generally. I construe it, instead, as a theory of practical reasons and the rules relevant to collecting and weighing practical reasons. I leave it as an open question whether such a theory is all there is to a theory of normativity or of value.

283 What does Korsgaard mean by the phrase “the maxim of acting on the impulse”? It means that the agent, if he believes that his practical identity demands that he (not) treat the impulse I as reason-giving in circumstances C, must make the same judgment about relevantly similar impulses in relevantly similar choice situations. For example, if an unwilling drug addict takes his practical identity to forbid him from treating a present desire for heroin as reason-giving because it would harm his kids, he must likewise think it forbids him from treating future desires for heroin as reason-giving so long as he is a father (227, 331).
personal meaning, and perhaps to a sense of pride or self-esteem. At another point, Korsgaard describes a practical identity as an “object of aspiration” (238). Is a connection to meaning, pride, or aspiration a necessary condition for having a practical identity? Or is any properly self-legislated principle of choice that is able to function as a practical guide a practical identity, regardless of whether it is also a source of meaning, self-esteem, or something to aspire to?

There is also a question about how a practical identity becomes normative for an agent. Korsgaard, following Kant, thinks that a principle of choice has to be self-legislated, and in a proper way, if it is to become a law for the agent. This means that it has to meet formal standards of volitional consistency: a self-inconsistent practical identity, for instance, can’t properly be willed: it will introduce inconsistency into her practical point of view. Does a practical identity also have to be self-legislated in a way that is consistent with Kant’s conception of personal autonomy? I would think so; but for reasons I’ll articulate, I’m not sure.

I’ll discuss these questions later. For the time being, with a rough sense of Korgaard’s proposal in mind, let’s see how an appeal to this concept is supposed to help.

Two Objections to Kant and Practical Identity

Why does Korsgaard think that Kant’s argument is unsound? Korsgaard acknowledges that Kant’s Categorical Imperative is subject to several empty formalism objections, and her own argument is an attempt to overcome at least two of them. The first empty formalism objection that motivates Korsgaard’s departure reflects her Cartesian way of construing the “problem” that confronts each reflective agent. Recall how Korsgaard thinks the general normative question arises. Reflection hears an internal voice—the call of conscience, the plea of a desire, the demand of an internalized taboo, or something else—and wonders whether the claim this voice
makes deserves consideration within practical deliberation. Conscience, the desire, the taboo, or something else is calling out to you, telling you what to do. Pure, primitive reflection—intellectual curiosity—leads you to wonder whether you should listen to it, and whether you should ultimately obey it. In my terminology, you want to know what practical reasons you have, and which are final.

The formal rules of rational willing—at least if they are taken to be a set of rules for volitional consistency—do not give the kind of practical advice the reflective agent “needs.” They tell you, for instance, not to pursue, at one and the same time, the satisfaction of both of two conflicting desires, and not to decide both to obey and to disobey the voice of, say, conscience. To do either of these things would be inconsistent. Hence the rules of volitional consistency do make demands, but they don’t tell you which desires generate practical reasons, or which practical reasons are final. Korsgaard concludes, if an agent is “to go forward” and be “active and rational,” she needs substantive principles for the collection and weighing of practical reasons. Without such standards, the reflective agent will suffer from stultifying disorientation in practical space, the practical analogue of Descartes’ experience at the end of *Meditations* I. Call this the “practical content objection.”

The second empty formalism objection is that if matter cannot be derived from form, then it follows that the matter of the Moral Law, the second and third formulations of Kant’s Categorical Imperative, cannot be derived merely from the thoroughly formal rules of volitional

---

284 The internal struggle is not necessary for the normative question to arise: if you find yourself with only a single desire, reflection might lead you to wonder whether you ought to lend it an ear, or do what it says.

285 This is overly simple. As argued in chapter 3, I think that there might be “factors” relevant to normative practical reasoning other than practical reasons. If a cherished personal ideal, for instance, includes norms that operate as defeaters, they are valid rules for the agent for whom the personal ideal is normative.
consistency. Let’s call this the “moral content objection.” Korsgaard construes this objection in a standard way. For a formal test to have substantive output, it is necessary for the input into the formal test to be substantive. Korsgaard, similar to Kant, thinks that principles of choice supply the substantive input into the relevant thoroughly formal test of practical reason. But Korsgaard does not accept weak universalism, which claims that an agent’s inclinations to act are “associated with” substantive principles for action. As evidence, think about how Korsgaard construes the problem of reflective agency. A clamorous desire does not provide the agent with a substantive principle for action, a principle which is a proper input into a test; on the contrary, the desire reveals that she “needs” a substantive principle of choice. This substantive principle of choice is not input into a test of normativity, it operates as this test.

Korsgaard thinks that practical and moral content objections are sound, if directed at Kant’s own argument. His argument not only fails to establish the universal authority thesis; it fails to give adequate practical advice to an agent who does not know which internal voices to treat as sources of legitimate practical advice. At the same time, Korsgaard has a theoretical burden similar to one of Kant’s. Following Kant, Korsgaard thinks that the rules of rational willing are thoroughly formal. Consequently, Korsgaard is compelled to seek out substantive principles of choice to which an agent is rationally committed, which will supply the substantive input into a thoroughly formal test of practical reason. This is where the appeal to an agent’s self-legislated normative self-conceptions enters her argument. Korsgaard seems to think that it is

---

286 Korsgaard argues that Kant’s universal law test, and so his first formulation, does have substantive implications. It is able, she thinks, to rule out the maxim that prescribes a lying promise and other maxims that prescribe actions that are instrumentally successful only if a social institution such as the institution of promise-making is in place (“Kant’s Formula of the Universal Law”). Someone might wonder, then, whether Korsgaard has changed her mind, given that in Sources she seems to accept the empty formalism objection that she rejects in this earlier paper. As I see it, in Sources, Korsgaard does not appeal to the universal law test in her argument. The thoroughly formal rules of rational willing that her argument begins from are not those contained in Kant’s universal law test. Her argument will take a different route.
more plausible to argue that an agent is rationally committed to his self-legislated normative self-conceptions than it is to argue that an agent is rationally committed to “maxims” that are “associated with” inclinations to act that happen to well up in the run of life. Perhaps so. If a practical identity is self-legislated in a more or less self-conscious way, if a self-conception is something from which an agent derives a sense of personal meaning or a sense of pride, or if it consists of one of his “aspirations,” it does seem plausible to argue that it constitutes a “genuine” practical commitment. Perhaps, then, it is a practical commitment from which an advocate of the Moral Law might attempt to leverage an agent to (greater) moral commitment.

In summary, Korsgaard appeals to the concept of a practical identity to overcome two empty formalism objections to Kant, and it suggests a two step process for reflective agents. An agent must self-legislate a practical identity, one or more, to escape the problem of reflective agency and its stultifying disorientation: “This is not contingent” (120) and “springs from” the reflective structure of the agent’s consciousness. Once the agent has self-legislated a practical identity, this normative self-conception, so long as it is self-legislated in a proper way, is a proper input into a thoroughly formal test of practical reason, which, Korsgaard promises, is able to compel—with the force of rationality—an agent to self-legislate “moral identity.” More later about how this test of practical reason is supposed to work. To foreshadow, it appeals to a transcendental principle.

My Dialectical Partner

Korsgaard’s appeal to the concept of a practical identity is of particular interest to me, given my general project. Her concept of a practical identity is similar in several ways to my...
concept of a personal ideal. In earlier chapters, I argued for the rational impotence thesis by saying that, in some cases, an agent’s personal ideals make up a part of the personal aspect of her will and include norms of rational impotence. Korsgaard agrees. But whereas I have suggested that these norms of rational impotence have subjectivist groundings, Korsgaard thinks that they become valid for the agent—are properly incorporated into the personal aspect of the will—only by acts of self-legislation that are not “determined” by anything external to the will, such as a motivational state. As a Kantian, she has the firm intention to stave off subjectivist proposals such as mine, not wanting to cede any ultimate normative authority to desire: ‘Because I want it’ or even ‘Because I care about it’ can’t be what grounds the act of will. What’s at stake is which theory does a better job of accounting for the normative authority of personal ideals and their associated volitional norms. Whereas I argued earlier, on one flank, against fellow subjectivists who deny the rational impotence thesis, I will presently confront an anti-subjectivist who appeals to ideas similar to mine.

My primary aim in this chapter is to argue that Korsgaard’s anti-subjectivist account of normative self-conceptions and their associated norms of rational impotence is subject to very serious objections. There are, I think it is fair to say, initial grounds for skepticism about Korsgaard’s proposal. First, it would seem that the reflecting self is able to ask, of a self-legislated practical identity, ‘Why treat its voice as normative for me?’ A practical identity is a practical standard, and it makes sense to ask whether the agent ought to follow its practical advice. There isn’t anything “incoherent, unnecessary, or impossible” about this question; an appeal to a practical identity does not seem to be a regress-ending answer to the general normative question. Nor is there anything incoherent, unnecessary, or impossible about asking whether the act of self-legislation of a practical identity is normative. Quite the contrary, as I’ll
discuss, there are times that it would be rationally advisable for an agent to reject a practical identity or to forego an act of self-legislation. The will could conceivably attempt to incorporate a practical identity that is internally consistent, but objectionable on other grounds.

Second, there are grounds for skepticism about Korsgaard’s ability to overcome the moral content objection. An agent could self-legislate a normative self-conception that resolves the problem of reflective agency but does not have any moral content, or even a practical identity that is positively immoral—one, say, which forbids her from treating the dictates of conscience as normative. How do thoroughly formal rules force such an agent to adopt, and to give priority to, a moral identity no matter what self-conceptions she possesses? How does it force her to “veer right” in the event that she is rationally committed to both a moral and an immoral identity? These are questions and objections to which Korsgaard will need to respond.

Korsgaard does attempt to respond to several of them, but I don’t think her responses are successful. My diagnosis is that Korsgaard’s commitment to the heteronomy thesis, an aspect of Kant that she does not give up, causes much of the trouble in her argument.

How to Evaluate Korsgaard’s Argument?

Before I begin my more thorough reconstruction and critical evaluation of Korsgaard’s appeal to normative self-conceptions, I have a few words, as a commentator. At times, Korsgaard herself downplays her differences with Kant, even important differences (98, 234-238). I intend to highlight them. I want to see whether Korsgaard is able to overcome the serious objections that plague Kant, and this project involves thinking about the ways in which her argument is

---

288 As I discussed at the end of chapter 4, though Korsgaard does not claim that the Moral Law’s demands are always overriding, she does think that the Moral Law’s dictates are important. She would not count her argument a success if it permitted people to treat moral reasons as exceedingly weak practical reasons.
different. Her theory is a Kantian constructivist theory, unswerving in its determination not to fall into subjectivism, modified to overcome several traditional objections to Kant. Do the modifications help?

To answer this question, it is useful to have a clear and distinct perception of what the newly-construed argument is. As Korsgaard herself admits, her line of reasoning is sketchy (91). It is clear what the two major steps in the argument are. But even at several key points in her argument, I am very puzzled about how to reconstruct it. Of course, I’ll simply have pull myself together and do my best to get the argument right. No use whining. But I won’t always succeed in imposing unity and integrity onto Korsgaard’s argument. It strikes me that the modifications that she makes to Kant’s theory lead into tensions, debilitating dilemmas even. Sometimes I’ll simply point these out, without attempting either to resolve them or to prove that she can’t.

For instance, Korsgaard’s theory of reflective agency seems to create an incoherent chain of command within the soul. She seems to grant the reflecting self the authority to call into question any purported source of practical advice and so to start regresses of ‘whys’. This is what leads to the problem of reflective agency. But she also grants to the legislating self the authority to create laws for the agent and so, it would seem, to end such regresses. What if the will self-legislates a practical identity and the reflecting self subsequently asks ‘why follow it?’? Which self has the final say? Is the principle of choice normative? Or does the reflecting self’s question reveal that it is “merely” another internal voice? Korsgaard, similar to Kant, emphasizes the

---

289 Gibbard makes similar remarks and admits that he isn’t sure if he’s gotten her argument right either. He calls Korsgaard’s book “a classic,” but “a frustrating classic.” He adds, modestly, “at least for me” (“Morality as Consistency”).
importance of “unity and integrity” to agency; so the presence of conflicting masters within the soul is a significant worry; it’s the kind of thing her theory is meant to stave off.

Korsgaard’s Departure from Kant

Here is one way of getting into the heart of Korsgaard’s family dispute with Kant. Korsgaard makes a distinction that Kant himself does not make. Kant takes the Categorical Imperative to be, at one and the same time, a rule of rational willing, a rule of personal autonomy, and the Moral Law—a trinitarian principle, a single principle with three purposes. Korsgaard, instead, regards the Categorical Imperative only as a rule of rational willing and of personal autonomy; she does not think that it is the Moral Law (98-100).

This isn’t a small modification. Korsgaard’s separation of the Categorical Imperative and the Moral Law has profound implications, which ramify through her argument. To mention one, in Korsgaard’s theory the will—the legislating self—is in an important sense independent from the Moral Law.\(^{290}\) Though Korsgaard continues to think of the will, following Kant, “as practical reason” so that its acts of self-legislation are subject to the formal constraints practical reason imposes,\(^ {291}\) she weakens those constraints considerably, relative to Kant. Strikingly, Korsgaard even claims that her Categorical Imperative does not rule out ethical egoism, the immoral maxim of the dear self in its most insistent mood. In Korsgaard’s theory, a person who properly self-legislates ethical egoism thereby comes to have the practical identity of an egoist and a “genuine

---

\(^{290}\) The ‘in an important sense’ is vague, but bear with me. Korsgaard doesn’t provide any language to be more precise. One of my questions will be what motivates Korsgaard to give to the will—which “is practical reason”—this level of freedom from moral constraint, especially given that she also wants to argue that practical reason does require moral behavior.

\(^{291}\) Korsgaard does not accept the language of imposition. She thinks that it suggests that these rules are “external” to the will, and imposed upon it by reason from the outside, whereas these rules are constitutive of practical reason and so the will, which “is practical reason” (236-237). I don’t mean to deny her point.
obligation”—an ‘ought’ of practical reason—to do as it says. The law of ethical egoism is a law for him, a genuine part of his self-constituting practical point of view. In accordance with her voluntarism, Korsgaard thinks that the will’s endorsement, in and of itself, and independent of moral considerations, enfranchises the dear self and its voice. Kant, of course, would say no such thing. He thinks that there is no way to properly legislate the egoistic maxim; the universal law test rules it out.292

One sense in which Korsgaard weakens Kant’s Categorical Imperative is that she does not include Kant’s universal law test within it.293 It is clear that Korsgaard regards Kant’s second and third formulations—to be respectful and fair—as parts of the Moral Law (99-100), but not as parts of the Categorical Imperative. Korsgaard regards what she calls “the universal law test” as a test of fairness: “What if everyone were to do what I am thinking of doing?” It is, that is to say, a part of the Moral Law. She doesn’t think, she is quick to point out, that this test “gives us the whole content of morality” (99); but she thinks “it is a mistake to think that it does not give us any content at all” (99). In other words, in Korsgaard’s view, the universal law test is a part of the correct normative ethical theory, a test of fairness; but there are moral considerations other than fairness. The upshot is that since the universal law test is a part of the Moral Law, it is something Korsgaard is arguing for, not something she argues from. Even more to the point, the universal law test simply falls out of her argument in Sources. Korsgaard says straight out that

292 Korsgaard indicates, though, that she does not share Kant’s anthropological view that the dear self is the most insistent, and psychologically most powerful, voice in each human heart (238-242).

293 As you read along in Korsgaard’s text, you could easily miss the point that Korsgaard does not include the universal law test within the Categorical Imperative. At one point she explicitly calls her Categorical Imperative the Formula of the Universal Law. She says, at one point, “Well, here is Kant’s answer. The categorical imperative, as represented by the Formula of the Universal Law, tells us to act only on a maxim which we could will to be a law” (98). But she doesn’t have the same conception of law as Kant.
she will leave for another day the job of trying to develop a normative ethical theory (100): “there are . . . proposals on the table about how to do that: John Rawls’ to name only one” (100).

Given her exclusion of the universal law test, Korsgaard’s argument for the universal authority thesis isn’t subject to the objections to weak and strong universalism. That’s the benefit she receives from her modification. She doesn’t have to attribute to an agent a maxim and its universal counterpoint to an agent simply in virtue of his acting upon a desire. And she doesn’t have to defend several of the more extravagant claims Kant makes about the logic of agency: that we are able to read off of an agent’s inclinations to act a general principle, a maxim, that makes no singular reference, or that each action has “eschatological implications,” indicating what kind of world you are consenting, or even aiming, to create. Setting aside weak universalism, she might seek out more plausible ways to argue that each person is rationally committed to substantive principles of choice that are proper input into her proposed test of practical reason, as such. (She does attempt to; she claims that (almost) everyone has a “jumble” of normative self-conceptions.

But this significant benefit comes with at least two significant costs. Here’s the first—I’ll mention the second in a moment. Given that Korsgaard intends to argue for the universal validity of the Moral Law, its right to tell each and every person how (not) to behave, she has to find some other route to her conclusion. She needs to appeal to other rules of rational willing to do the work for her that the universal law test was to do for Kant. We should ask, of course, whether the rules

294 Or, at least, these objections would apply to her normative ethical theory, but not to her argument for the universal authority thesis.

295 See, also, her further remarks about the logic of agency, what willing an action does and doesn’t commit an agent to (227).

296 A reflective creature that doesn’t has a different problem: no way to go forward. He is able to escape her arguments, but at the cost, Korsgaard thinks, of giving up the possibility for agency.
of rational willing Korsgaard appeals to in the second step of her argument are more plausible than Kant’s. I don’t think they are.

So, Korsgaard reconstrues the Categorical Imperative. What does hers require? She tells us that the Categorical Imperative is the rule “adopt a law, any law.” This rule is created, in part, by the problem of reflective agency; without a practical point of view from which to evaluate desires a person does not have any way to make principled decisions.\(^{297}\) The reflective agent has a need, the Categorical Imperative is rule that tells him to meet it.\(^{298}\) Hence, the Categorical Imperative itself doesn’t adjudicate between conflicting desires, and it doesn’t accept or reject a desire of the moment; it tells the agent to adopt—to self-legislate—substantive principles of choice that are to do that work. The Categorical Imperative is a principle that demands that an agent create a practical point of view.

Once this command is successfully obeyed, Korsgaard believes, the agent has a practical point of view from which to evaluate his desires, as well as the norms of his community, the dictates of God and politicians, or any other purported source of legitimate practical advice. This is the sense in which Korsgaard replaces the universal law test with the normative self-conceptions that an agent has properly self-legislated. No doubt, an agent will often have hard work to do in deliberation; the ethical egoist has to figure out what really would be in his best,

\(^{297}\) The Categorical Imperative also reflects the proposals Korgaard rejects. It is because she thinks an agent needs a practical point of view, and she won’t treat, among other things, desires of the heart as the proper source of such advice, that she affirms this Categorical Imperative.

\(^{298}\) Do needs always generate rules to meet them? If so, how? A subjectivist is going to say that if a need, say a biological need for replenishing fluid, causes a desire, thirst, the agent will have a practical reason to meet the need. Also, if the need is connected to the person’s health, then a desire for health (motivated and/or basic) generates a practical reason to drink. How would Korsgaard handle these cases? Generally, I’m not sure. But with respect to the “need” created by reflective agency, she thinks that it supplies you with a rule—the Categorical Imperative—only if you “treat this part of your identity as normative.” This strikes me as an implausible chain of reasoning. Do we need, likewise, to treat our biology as normative, by way of an act of self-legislation, for biological needs to play a role in giving us reasons?
long-term self-interest. But he does not suffer from stultifying disorientation; his practical identity often gives him practical advice about how to make practical decisions.

Another aspect of Korsgaard’s modification, closely related to what I’ve already mentioned, is that she weakens Kant’s conditions for lawfulness. She regards the Categorical Imperative as a complex principle of volitional consistency, but it is not a universalizability test for maxims. As mentioned, Korsgaard thinks that the law of an ethical egoist fulfills the Categorical Imperative’s demands, which Kant would never admit.

This brings us to the second theoretical cost Korsgaard incurs from weakening the Categorical Imperative. Kant himself says that his theory of personal autonomy “robs” the will of any substance; it calls, among other things, any desire, any social rule, any political decree, and any dictate of God “alien” or “external” to the will. He attempts to overcome the empty self objection by claiming that our inclinations to act, the raw material that presents itself to the will, are associated with principles of choice, maxims, and that the universal law test takes these maxims as proper input. As discussed in previous chapters, he thinks this test has substantive moral implications. Since Korsgaard does not appeal to the universal law test, she “robs” the will of the one standard—the one test—that Kant includes in it. This appears to leave the “legislating self,” in Korsgaard’s theory, without any grounds for making a choice, in any meaningful sense of ‘choice’, about which maxims to self-legislate, and without any clear sense of where the relevant maxims that are to be tested are supposed to derive from.

\[299\] That’s a little loose. Both Kant and Korsgaard accept an instrumental principle. But the instrumental principle only operates if a rationally valid end is present. The mechanisms of instrumental reasoning need to be supplied with an end for which they are able to seek a means. Kant, at least, indicates how an agent’s desired ends play a role, once put through the relevant test, in generating practical reasons.
Think of it this way. In a particular choice situation, the legislating self could make a substantive practical decision about whether and which practical identity to self-legislate on the basis of another practical identity that has been previously self-legislated. But this appeal only presses the question back one step: how did the will choose this other practical identity? Korsgaard’s commitment to the heteronomy thesis seems to say that a practical identity can’t be chosen merely on the grounds ‘I care about being a good such and such’ or ‘My minister tells me I should be a such and such’. In either case, such a decision treats something “alien to the will” as the source of a practical reason, a practical reason to adopt a practical identity. But such things are only permitted, in the Kantian theory, to “influence” but not “determine” the will.

In summary, Korsgaard’s Cartesian starting points indicate that the reflective agent is desperately in need of substantive principles of choice; he suffers from stultifying disorientation in practical space without them. Her theory demands that the will make a choice, but seems to leave the will with no proper grounds for choosing any. If Korsgaard sticks to her commitment to the heteronomy thesis, it seems that she’ll have to admit that the will’s “first” act of self-legislation has to be nothing more than a mere “whim of the will,” an arbitrary plunking for some principle of choice or other from among the myriad of options. But this seems problematic. Why not prefer a theory that permits an agent to make practical decisions about which personal ideals to treat as normative by determining, best he can, which best cohere with the desires of his heart? If a father were to adopt a principle such as ‘Be patient’ on the grounds that it will help him satisfy his desire to be a decent parent, wouldn’t that count as a choice, in a meaningful sense?

Would Korsgaard ally herself with existentialists who believe in “radical freedom”? If so, see Charles Taylor (“Responsibility for Self”).
A Word from Kant

Kant could cry foul, of course—at Korsgaard. He might point out that it is unfair—a straw man fallacy—to gut his Categorical Imperative and then to describe it as empty. But I think Korsgaard could easily defend herself against this accusation. There are serious objections to weak and strong universalism, such as those presented in the previous chapter. I am willing to acknowledge that Korsgaard’s departures from Kant could reflect a self-conscious choice, grounded in a sensitivity to these objections—in which case she’s not misconstruing Kant; she’s giving up implausible aspects of his argument.

Whatever the case may be, Korsgaard herself weakens Kant’s Categorical Imperative. I’ve raised some initial objections, but let’s make our way into the heart of Korsgaard’s argument to see if hers does better.

Two Stages in Korsgaard’s Argument

Korsgaard’s argument is made, broadly speaking, in two steps (91). First, she argues from the nature and conditions of agency to her Categorical Imperative; second, from her Categorical Imperative to the Moral Law. Korsgaard’s broad argument mimics Kant’s. But it differs in important details. Let’s take a look at the some of the relevant details.

The Argument for the Categorical Imperative

In the first stage of her argument, Korsgaard indicates what her Categorical Imperative is and what it demands of us. As mentioned, Korsgaard regards the Categorical Imperative as a rule of rational willing and a rule of personal autonomy; it expresses a set of preconditions for rational
and autonomous behavior, a (complex) rule for self-creation: ‘Self-legislate at least one practical identity’. Her Categorical Imperative is not a test for substantive principles of choice, which takes them as input; it is the rule that we self-legislate some, at least one.\footnote{The difference between Kant and Korsgaard could also be put in terms of freedom. Kant means for the Categorical Imperative to be a principle of self-governance; an agent is personally autonomous only if she self-legislates and acts upon maxims that pass the universal law test. The type of unfreedom it is intended to stave off is heteronomy. A fully rational agent has self-legislated principles of choice that thereby become a part of the personal aspect of her will. When she acts in accordance with her will, she is at the helm. She isn’t a “slave” to her passions, or to any other type of alien influence. Also, Kant believes that, if followed, the Categorical Imperative, given that it is, he thinks, the practical principle of non-contradiction, will unify an agent’s practical point of view; there will be, at least, no conflicts between the rational and personal aspects of the agent’s will. There won’t be any inconsistency within her will. The rules of rational willing, many of which are rules of volitional consistency, impose integrity upon both her, and her subsequent volitional activity. Korsgaard means for the Categorical Imperative to be a principle of self-creation. Action, she thinks, is not possible without having a practical point of view. The type of unfreedom it is intended to stave off is complete disorientation in practical space. Indecision is stultifying. If you do not have a practical vantage point from which to make practical decisions, there is no way for you to make principled choices: “the reflective mind . . . cannot commit itself or go forward” (93). This is because, without a practical point of view, there is, as Korsgaard puts it, “no ‘you’” to choose one and not another option. Korsgaard demands that the agent adopt principles of choice by acts of will that can then be used to stand in judgment of desires.}

Korsgaard consistently expresses the idea that if you are to be an agent, and not merely an arena in which desires struggle for satisfaction, it is necessary for you to have (in my terminology) a practical point of view, a self-constituting vantage point from which to make principled decisions and to govern your conduct: “This is not contingent” (120), and

What is not contingent is that you must be governed by some conception of your practical identity. For unless you are committed to some conception of your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another—and with it, your grip on having any reason to live and act at all. (121).

Frankfurt and I share with Korsgaard the idea that self-governance requires a practical point of view. But we construe it differently. First, it’s not clear why an agent who seeks to satisfy a desired end can’t have a “grip” on himself as “having [a] reason to do one thing rather than another.” He has a reason to take the effective means to his ends, not to ignore them. (It’s only if
you accept the heteronomy thesis, or something similar, that you’ll think otherwise.) Second, since Korsgaard accepts Kant’s claim that practical reasons are derived from principles of choice, she thinks that a practical point of view consists entirely of principles of choice (98)—not, that is to say, of ruling passions or desires of the heart. In and of itself, your will does not include any substantive principles of choice; the rules constitutive of the rational aspect of your will are thoroughly formal. And so, Korsgaard thinks it is necessary—a precondition for rational and self-governing behavior—for you to incorporate substantive principles of choice into your personal will, which otherwise would be thoroughly empty and utterly silent when desires, or whatever, clamor for attention. You have to create your personal will, by acts of will. (This is how Korsgaard articulates both her voluntarism and her constructivism.)

The first thing that the Categorical Imperative demands is that you incorporate at least one principle of choice into your personal will. In this way, Korsgaard tells us, obedience to the Categorical Imperative “brings existence” to the self that is consequently able to make rational or principled choices. Once you have a practical law, you are able to make practical decisions on the basis of it. Korsgaard also thinks that it is necessary to have a substantive practical point of view if we are to exercise personal autonomy. Without a substantive practical point of view, there is no “you”—no self, in the relevant sense—who is able to govern. Notice and keep in mind the link between the Categorical Imperative and personal autonomy.

At a general level, we might describe the Categorical Imperative as a two-fold rule for maxim-legislation. The first demand it makes is that you must self-legislate at least one principle of choice into your personal will. Korsgaard’s claim is even stronger: no practical reasons exist for you unless and until you self-legislate at least one principle of choice (125). It’s misleading, I think, when Korsgaard says “The categorical imperative . . . simply tells us to be autonomous” (220). What it tells us to do is to meet a precondition for personal autonomy: to construct a self that is capable of governing.
of choice, the second demand is that any principle of choice you self-legislate has to be legislated in the proper way—“there are right and wrong ways” to create a practical point of view (35). A candidate maxim has to meet several formal conditions. This test, whatever it is, is a formal test, and so she describes her theory as a form of procedural realism.

Korsgaard doesn’t say much about these formal requirements. Since Korsgaard does not include a universalizability test within the Categorical Imperative, the test doesn’t include the thought experiment associated with the universal law test. Korsgaard does tell us, though, that the function of these formal requirements is to “impose unity and integrity” on the self (229)—in my terminology, on the agent’s practical point of view. This suggests a similarity to Kant and his principles of volitional consistency. Accordingly, I suspect Korsgaard regards the Categorical Imperative as a rule that forbids the willing of an internally consistent maxim and the willing of mutually inconsistent maxims. Korsgaard also seems to think that a maxim is properly self-legislated only if it is able to resolve the “problem” that confronts the reflective agent; it has to be able to save him from stultifying disorientation. Perhaps, then, a principle of choice that is so vague that it can’t issue in useful practical advice would be ruled out. Ethical egoism, Korsgaard must be thinking, provides the relevant kind of practical guidance. It tells an agent to screw himself to the task of maximally benefiting himself; it advises him, for instance, to suppress whatever fellow-feeling he might have whenever doing so would maximize his long-term benefits.

Perhaps to drive home the point that the formal requirements of the Categorical Imperative are minimal and thoroughly formal, Korsgaard mentions that the requirements of her Categorical Imperative are met even by self-legislating “the law of a principled wanton.” This

---

304 Well, then again, there is the paradox of happiness.
law tells you simply to let the strongest desire win, without participating in the struggle between desires that ultimately leads to behavior. It might seem as though principled wantonism isn’t a principle that gives practical advice—it doesn’t tell you which of two conflicting desires to favor, for instance. But a principled wanton has a program, a program of non-intervention. Apparently, that’s enough to meet the demands of the Categorical Imperative. The principled wanton has a “thin” self but one “thick” enough to exercise rational agency—it takes work, I suppose, not to intervene in the struggles between your desires. (Similar to a mother who has a hard time letting her children resolve their own disputes, she has to resist the urge to enter the fray.)

The formal constraints Korsgaard appears to include within the Categorical Imperative are intended to be, I think, relatively weak and relatively uncontroversial. The strongest claim that Korsgaard explicitly includes is a principle of consistency in judgment. Any principle of choice that is adopted “essentially involves a reference to other occasions on which I might do otherwise” (231). But she even qualifies this. Consistency of judgment is required of an agent only relative to his self-legislated normative self-conceptions. To self-legislate a principle of choice rationally commits you to obey it unless and until you change your mind. In the meanwhile you have to treat its practical advice as practical advice for you and conform your behavior and volition accordingly. This set of claims combines prescriptivity and consistency in judgment: you must make similar judgments in similar choice situations when under the direction of the same practical identity.

According to my interpretation, Korsgaard thinks that you are able to make practical decisions without committing yourself to acting in the same way in any relevantly similar set of circumstances (227). But if you never make a stable commitment to a set of principles of choice—with at least one member—then you will not be able to identify yourself as an agent. You won’t have a stable practical point of view from which to make decisions.
If the reflective agent has prompted the will to make a law for itself, and he does so properly, meeting these conditions, then he is able “to go forward” and to be “active and rational.” Korsgaard leaves several important questions unanswered, though. Does the Categorical Imperative require a maxim to lack singular reference? She doesn’t say. If she does, she needs to argue for this. It would seem that a principle of choice that includes singular reference could resolve the problem the reflective agent confronts. Paul’s father is able, I imagine, to see himself as active and rational when he is following the principle ‘To benefit Paul, help him with his homework in the evening’. This principle gives him substantive practical advice about what to do: ‘Don’t spend the entire evening at the bar’, for instance. The principle also tells him how to treat various desires; if he finds himself wanting to watch television the whole evening, his principle tells him to call that desire a temptation and to reject it, at least as giving him a final practical reason.

A caveat: I said Korsgaard expects an agent’s self-legislated maxims to be mutually consistent. Surprisingly, at one point she claims she doesn’t expect this, at least not invariably. Korsgaard tells us that it is part of what makes “life hard” that we sometimes confront “unconditional” obligations that conflict (126). Given her voluntaristic constructivism, she thinks a principle of choice becomes a practical law only if self-legislated. It seems to follow that it is permissible, at least in some cases, to will principles of choice that conflict, even in deep ways. I suspect that Korsgaard thinks that some self-legislative acts are to be ruled irrational because they introduce inconsistency into the personal aspect of the will. A person who self-legislates the principle ‘Be charitable’ and ‘Keep everything you own to yourself’ is guilty of a rather obvious and dumb inconsistency, and he will not be able to “go forward” without violating one or other of his principles. But there are more compelling cases. We could interpret Agamemnon at Aulis as
suffering from a more understandable and defensible volitional inconsistency; he has made unconditional commitments to both his daughter’s and his army’s welfare. Due to his great misfortune, he is forced to choose between them. Think of the parent with multiple children. It is arguable that a parent might want to go on holding unconditional commitments to lives of his various children even if he notices that he could conceivably be placed in a situation in which he is forced to choose between them.

I mention this only because it will create a difficulty for Korsgaard later in her argument when she seems to argue that an agent with an immoral identity should give it up in favor of moral identity. Presumably, at that point she will argue that the immoral agent has to give up one of his commitments by appealing to a practical principle of non-contradiction to argue that it is practically irrational to will (at least self-consciously) inconsistent maxims. If there are exceptions to the practical principle of non-contradiction, then how do we know that this contradiction is to eschewed rather than grasped? If the immoral identity is unconditional, and very important to the agent, couldn’t he—similar to Agamemnon—will both identities? Maybe Korsgaard has some way of fitting these ideas together. It’s hard to tell, though, since she doesn’t speak to the issue.306

In any case, Korsgaard emphasizes that her Categorical Imperative is thoroughly formal: it does not rule out candidate maxims on the grounds of their content. It permits you to will

---

306 Does Korsgaard reject the practical principle of non-contradiction? She does think that there are instances in which two volitional states contradict: “Conflicting obligations can both be unconditional; that’s just one of the ways in which human life is hard” (126). Since she thinks obligations derive from acts of self-legislation, this seems to imply that she is willing to admit that contradictory acts of self-legislation are not rationally forbidden. To be charitable, though, this commits her only to saying that the practical principle of non-contradiction is not absolute; there are instances in which it is permissible to violate it. She never tells us, though, when this is, and when it isn’t, permissible. She needs to, though. Korsgaard argues that the ethical egoist ought to repudiate his ethical egoism in order to have a consistent practical point of view. If it is rationally permissible to have conflicting obligations, why can’t he?
ethical egoism, if that happens to be your cup of tea, or Quakerism, or being a good brother to your younger sister. We might notice the set of all principles of choice that meet these requirements is named ‘legion’. But that doesn’t seem to trouble Korsgaard. She says ‘Adopt one’ and ‘Make sure it is able to do the work that you need it to’.

A (Seemingly) Central Aspect of Korsgaard’s Answers, at a First Approximation

To respond to the practical and moral content objections, Korsgaard appeals to the concept of a “practical identity,” a normative self-conception, and what she takes to be its characteristic role in practical deliberation. It seems curious to me that it is possible to describe the contours of the first step in Korsgaard’s argument without any clear mention of the concept of a practical identity. (Do the ethical egoist and the principled wanton necessarily find meaning or a sense of pride in their respective self-conceptions?) The Categorical Imperative is construed in terms of the self-legislation of principles of choice, though Korsgaard seems to treat the concept of a practical identity as an essential aspect of her argument. Perhaps this means that a principle of choice, any principle of choice, that is self-legislated thereby becomes a valid practical identity.

But Korsgaard could easily be interpreted as building more into what a practical identity is. She describes a practical identity as a self-conception that is connected to self-esteem and to personal meaning. This is suggested, at least, by the gloss of ‘practical identity’ Korsgaard provides (101), and her description of a practical identity as an “object of aspiration” (238). These remarks suggest that Korsgaard has in mind more than the self-legislation of a “bare”

---

307 This gloss seems to mimic Kant’s idea that an agent must act “under the idea of freedom.” Is this important? For instance, does Korsgaard mean to suggest that it is a necessary condition upon action that we act “under the idea” that we are a particular kind of person, with a particular practical identity?
principle of choice. It seems possible, after all, to self-legislate a principle of choice that is not connected to self-esteem, a sense of meaning, or any “aspirations.” I might decide to adopt the principle ‘Watch television for a half an hour each evening before going to bed’ with the idea that this will help to relax me. My obedience to this principles demands will hardly make me proud, give my life meaning, or make me believe I’ve accomplished something with my life.

Korsgaard also, at times, seems to regard a practical identity, as a normative self-conception with a striking “scope.” Korsgaard says that to adopt a practical identity is to adopt “a way of life and a set of projects, and the new desires which this brings in its wake” (239). This remark fits with Korsgaard’s remark that a practical identity is an idea “under which” you find that your life is worth living. But this further distinguishes a practical identity from a (self-legislated) “bare principle of choice”: there are instances in which an agent adopts a principle of choice that is intended to be temporary or that is intended to govern only a relatively unimportant part of her life.

Korsgaard could choose to distinguish, in any of these several ways, between bare principles of choice and practical identities. But then she needs to argue why the self-legislation of a practical identity, as opposed to a bare principle of choice, is a precondition of rational agency. It seems a bare principle of choice could solve the problem of agency. If all that the reflective agent needs is substantive advice that is sufficiently determinate to issue practical advice about his choice situations, a bare principle will do. An agent needn’t find a practical standard personally meaningful to be able to figure out what practical advice it would give.

---

308 Ethical egoism is a principle of choice, but is it, by necessity, a practical identity? I would think some ethical egoists, those who are pure hedonists, do not care about meaning; they want to maximize pleasure.
Korsgaard could also choose to say that many of her descriptions of normative self-conceptions are common, but not essential features of having a practical identity. Self-legislating a bare principle of choice is enough. That raises several questions, though. Is the appeal to the concept of a practical identity otiose? Second, is the mere self-legislation of a bare principle of choice the type of practical commitment that implicates an agent in valuing and so can get the second part of the argument going? If I don’t much care about a principle of choice I’ve self-legislated, if it isn’t meaningful to me or a source of pride, perhaps I’d be willing to give it up if I saw that it had serious practical implications.

Several objections apply. First, Korsgaard, if she means to claim that a practical identity is more than a bare principle of choice, has to argue that more than the adoption of a bare principle of choice is necessary for the resolution of the problem of reflective agency. And the more that Korsgaard builds into the conditions for having a practical identity, the less plausible it seems that this is necessary for reflective agency.

Second, I think Korsgaard’s talk of the “need” for principles of choice is already problematic. The heteronomy thesis, along with Korsgaard’s rejection of Kant’s universal law test, manufactures, it seems to me, a false crisis. A subjectivist theory has some very plausible things to say about how to escape stultifying disorientation and how to acquire meaning. First, subjectivism doesn’t really let the regress get going; it treats an agent’s basic motivational states as rationally potent, and so it finds the “problem” a false problem. It is rationally advisable for agents, given what most of us care about, not to be without any pursuits. Korsgaard is correct that if a creature with the capacity for reflection could manage to find anything to sensibly treat as normative, he would be left disoriented in practical space. So, there is a very general “need” for a source of practical reasons or practical advice. But Harry Frankfurt has argued plausibly for a
very general account of why we ought to care about something: we hate boredom and so we need useful work, something to keep us preoccupied. Similar to Korsgaard, at least in the first stage of her argument, he thinks there are, as a matter of the logic of agency, a myriad pursuits that could fulfill this need. We need things to love, to throw ourselves into, with confidence and a sense of direction. What we love, and what we are capable of loving, is a contingent matter, which is partly the consequences of nature, personal history, social enculturation, and, perhaps, decisions the person himself has made.

Korsgaard wants to construe the ‘need’ that undergirds the Categorical Imperative in a non-subjectivist way: this “need” arises, not from a desire, but from “the structure of a reflective mind,” and so from an inescapable aspect of your identity. I’ll evaluate this in a later section.

Korsgaard thinks her appeal to the concept of a practical identity and its characteristic role in practical deliberation answers the practical content objection. A practical identity contains substantive rules for volition and behavior; and if it is properly self-legislated, it supplies valid practical advice to the agent.

Objections to Korsgaard’s Argument for the Categorical Imperative

First, as mentioned earlier, an appeal to a practical identity is not a regress-ending answer to the general normative question, and so a self-legislated practical identity cannot be an ultimate source of practical reasons, as Korsgaard seems to think. An appeal to a practical identity provokes, legitimately, another ‘why?’ from restless reflection. A desire of the moment isn’t the only thing that might clamor for attention from the will. From the perspective of the reflecting self, an agent’s practical identity is simply another voice within the agent’s internal dialogue, without any clear claim to special authority for practical deliberation. By her own standards,
Korsgaard should say that the practice of treating a practical identity as a source of practical reasons is a type of practical deference that is in need of practical justification. Asking ‘Why should I treat my practical identity as a source of practical reasons?’ is not ‘incoherent, unnecessary, or impossible’ (33). Far from it; there are occasions in which it is rationally advisable to get rid of one. It is possible to imagine a practical identity that has significant psychological influence upon an agent, largely harmful. For instance, imagine a person with little athletic skill and no athletic promise who adopts the practical identity of a jock, and acts accordingly. He goes to the gym to work out day in and day out. And along with the aches and pains, insult is added to injury. He suffers the derision of the athletes his identity compels him to surround himself with, and from the personal shame that comes with the failures that his native endowment make inevitable. A friend asks him, “Why put yourself through all of that?” The answer “I regard myself as an athlete” simply provokes the question whether it is a good idea to have that self-conception. This question is not “incoherent, impossible, or unnecessary”; in fact, given the psychic costs of the identity, it seems best to severely restrict, if not reject, the identity. A practical identity comes with costs, and it makes sense to ask whether those costs are worth incurring, all things considered.

It is also possible to ask other types of practical-standard-based questions about a practical identity. “Is my practical identity ignoble, or small-minded, or parochial?” More generally, it is not always a good idea to treat a practical identity as a source of practical reasons and practical advice. Again, it seems the answer will sometimes be ‘yes’. Far from ending a

---

309 Also, mere practical deference to a practical identity is heteronomous. And someone who defers to a practical identity that, say, he acquires from his upbringing lacks the type of personal freedom that Kant emphasizes. One of the strengths of Kant’s own theory is his compelling vision of a robust personal autonomy. Myself, though I am not inclined to think that practical rationality itself requires that we attempt to answer all of the questions that restless reflection coherently asks, but I don’t doubt that the ideal type of personal autonomy that Kant emphasizes would require an agent to do so.
recess of whys, an appeal to a practical identity generates a wide range of new questions. For that matter, Korsgaard seems to think the question ‘why?’ makes sense even if it is motivated by nothing more than intellectual curiosity or a spasm of nervousness. What wouldn’t it make sense to ask ‘why’ about?

At one point, Korsgaard seems to deny these sensible thoughts. She calls the will’s authority “beyond question” (104). If Korsgaard intends to overcome this type of objection by appealing to the mere fact that your will—your legislating self—happens to have legislated the self-conception, this doesn’t help either.\(^{310}\) Acts of self-legislation admit of reflection; the reflecting self can step back from them and take them as objects for scrutiny. And as I see it, an agent often has good reason to scrutinize his own acts of will. An agent’s volitional activity might reflect ignorance or carelessness, or worse—an agent might make practical decisions in a fit of self-loathing. Why think that the will’s activity—whether with respect to self-conceptions, or more generally—always “creates” valid practical laws for an agent? If the agent wants to take care, he should scrutinize his acts of self-legislation, too.\(^{311}\)

To escape this set of objections, it would seem that Korsgaard is forced to posit, in addition to a test of reflection for desires, a test of reflection for self-conceptions and/or for the act of self-legislating a practical identity. Self-conceptions, and the act of self-legislating them, don’t seem to get us to the bottom of things, to the unconditioned. I suppose, in a sense, Korsgaard agrees. Given her response to the first empty formalism objection, the moral content

\(^{310}\) Also, why can’t the will legislate something other than a practical identity?

\(^{311}\) Realists should be aggrieved by Korsgaard’s way of ending the regress. Their theories were rejected on the grounds that they merely express confidence that there really are objective values. Korsgaard responded by saying that this appeal to a purported regress-ending objective standard doesn’t “speak to” the reflective agent in his time of need, “he doesn’t see it just now.” Is the claim that the will has absolute authority to self-legislate a practical identity an obvious and undeniable truth? I don’t see it right now.
objection, she should agree that the reflecting self is able to call into question an agent’s practical identity. For at that point in her argument, Korsgaard claims to think that action-guiding self-conceptions do admit of reflection and its questions. It is, she says, reflection on the substantive action-guiding self-conceptions that we have self-legislated that is supposed to lead us to recognize the validity of the Moral Law. It is when we step back from our contingent values that we realize what we have in common with others, and what capacity—the capacity for reflection—makes our having values possible. This suggests that she is treating the act of self-legislation of a practical identity as though it were not regress-ending: “It befits an adult human being,” she says, “to question” (248), and it is a part of an important type of existential “maturity” to recognize the contingency of our various self-conceptions (248), and to begin to seek an answer that is unconditioned.

But if a practical identity admits of a request for practical justification, how does an appeal to it resolve the “problem” of reflective agency? As we’ll discuss later, Korsgaard says that your will—your “legislating self”—has the practical authority to make laws for you. Whatever your self-legislated practical identity tells you to do is, for you, an obligation, and the will’s authority is “beyond question” (104). She explicitly eschews the claim that, for a person with a practical identity, she merely “experiences” its demands as obligations. They really are. (You can ask for practical justification, she would admit, but the answer ‘yes, it has it—you willed it yourself’.) This suggests that Korsgaard is treating an act of self-legislation of a practical identity as regress-ending.

This is one place I think Korsgaard’s theory pulls apart. Either a self-legislated practical identity is regress-ending, or it isn’t. She seems to want the reflecting self to acquiesce to contingent self-conceptions when she speaks of the Categorical Imperative, but to call them into
question when she is arguing for the validity of the Moral Law. This is what provoked me to say, at the beginning of the chapter, that there seems to be a conflict within the mind’s “command structure.” The reflecting self has the right to demand a satisfactory, regress-ending answer to the normative question, asked about conscience or any other internal voice. Since the will is independent of the Moral Law, Korsgaard seems to attribute to it the practical authority to enact practical laws “as it pleases.” Does reflection have the right to demand the will to re-think its acts of self-legislation? Or does the will have jurisdiction over the reflecting self, with the authority to shut it down? Korsgaard seems, at various stages in her argument, to answer ‘yes’ to both questions.

There is an analogous tension between the two ways in which Korsgaard construes our experience as reflective agents. Given our capacity for reflection, we are able to call into question any mental state or volitional tendency; hence the possibility of stultifying disorientation: where in the world do we stand when we ask these questions? But Korsgaard also wants to say that we all confront, in conscious experience, the personal obligations that derive from our self-legislated self-conceptions. This confrontation is what leads us to the “fact of value”: we can’t help but think, most of us, that there “really are” legitimate values, which exert a legitimate influence upon us (4). Which is it? Either you and I confront the problem of reflective agency, and are afflicted with a need for an unconditional answer to it, or we feel the binding force of the obligations to which our self-legislated practical identities give rise, and can rest assured in them. How can it be both?

It seems to me that Descartes has the upper hand. The stronger answer is that our capacity for reflection is able to call into question our action-guiding self-conceptions, and the acts of will that endorse them.
Let’s say my judgment is correct: a self-legislated practical identity does admit of reflection and its questions. This, it seems to me, is what Korsgaard needs to say if she is to get the second stage of her argument off the ground. Otherwise, you and I could simply rest assure in our contingent self-conceptions. They give us, it seems, genuine practical reasons and real obligations. Korsgaard claims we can’t, not even in the ones we’ve self-legislated. Restless reflection might call these into question, too. What she says is that we ‘ought’ to take them as objects of reflection; we “should never stop reflecting until we have reached a satisfactory answer” (257-258).

Moral Identity is But One Possible Practical Identity

Here are some initial objections to the second part of Korsgaard’s argument. First, Korsgaard herself includes moral identity (the identity of a citizen in the republic/kingdom of ends) in a long list of possible action-guiding self-conceptions from which agents are able to choose. Moral identity seems to be nothing more than one possible identity among many others. If a practicing agent, qua agent, must have a principle of choice, there are many other principles of choice he might choose. So, even if Korsgaard’s argument for the requirement to have a practical identity is successful, it does not entail that an agent must have moral identity. At first glance, there is not anything, so to speak, “rationally inescapable” or otherwise special about moral identity; it is but one option among many.

And so, even if Korsgaard were able to argue that the Categorical Imperative, on her construal, is able to compel an agent to adopt a substantive action-guiding principle for action, so that the Categorical Imperative is minimally demanding, it seems that neither the little categorical
imperatives, nor the Formula of the End in Itself, nor the Formula of the Kingdom of Ends is on the cusp of being derived.

Also, since—as I argued in the last section—a practical identity is not a regress-ending answer to the general normative question, then an appeal to moral identity isn’t either. Even if an agent were to adopt moral identity, this would not end the regress of whys. Presumably, moral identity has practical costs—being a good citizen entails personal sacrifices—and this fact motivate the question whether the agent ought to continue to self-legislate moral identity.

The Mafioso and the Obligation to Assassinate

Second, what if an agent self-legislates an immoral identity? G.E. Cohen has us consider a particular mafioso, whose commitment to his practical identity is whole-hearted—he doesn’t feel moral qualms about his mafia activity—and has a heroic sensibility. Cohen argues that Korsgaard’s theory of practical reasons commits her to the claim that the mafioso’s practical identity—despite its immorality—is for him a source of practical reasons—obligations, in Korsgaard’s view—to kill people. Far from securing the practical authority of respect for persons, Korsgaard’s appeal to the concept of a practical identity suggests that a person’s existential commitments, no matter their content, “create obligations” for her.

Two More Differences with Kant

Korsgaard deviates from Kant in two other ways that are relevant to my discussion. The first has to do with the character of practical reasons. Korsgaard says, of her weakening of the

Categorical Imperative, that it implies “a deep element of relativism in the system” (113). I am not sure ‘relativism’ is an apt term. Korsgaard isn’t a moral relativist. She thinks the Moral Law is universal, it has the same content for everyone: it expects everyone—the same as Kant’s second and third formulations—to be respectful and fair.\(^{313}\) Her theory is “relativistic” only with respect to practical reasons: an agent’s practical reasons are “relative to” his properly self-legislated principles of choice. But Kant thinks the same thing. So this “relativity thesis” doesn’t really distinguish her from Kant.

What centrally distinguishes Korsgaard from Kant, as I’ve been saying, is that her Categorical Imperative lacks Kant’s universal law test, and the will has the practical authority to make laws free from moral constraints. Korsgaard admits with equanimity that there is nothing within the Categorical Imperative itself to force an agent to coordinate his maxims with anyone else’s needs or personal projects. Immoral behavior, Korsgaard thinks, can be done on the basis of “genuine” practical reasons (257). That’s the claim Kant doesn’t accept.

\(^{313}\text{We should be clear about what Korsgaard takes the Moral Law to be. Given that she changes the content of the Categorical Imperative, the reader might wonder whether she has changed the content of the Moral Law, too. The answer is ‘yes, but not much’. It is relatively easy to describe, at least at a general level, how Korsgaard differs from Kant with respect to the content of the Moral Law. When Korsgaard speaks of the Moral Law, she means to refer only to what Kant calls the matter of the Moral Law, and she tends to favor something very similar to Kant’s Formula of the Kingdom of Ends, his third formulation. But she makes two changes. First, Korsgaard speaks at times of a “republic” instead of a “kingdom” of ends. Kant speaks of the moral community as having a king—God, the supreme lawmaker. Korsgaard does not. In her view, everyone within the moral community is equal, and there is no supreme lawmaker. The social rules that ultimately ought to govern the moral community—the community of ends, and each of its citizen—must arise from the general will, and not from the will of a single, supreme legislator, the king. Korsgaard explicitly remarks that the Moral Law is not an empty formalism (99), but she doesn’t intend in Sources to specify very much of the content of these rules for society. She simply expresses confidence that there is “some way of determining what those laws would be” (100). And her remarks reveal that she gives the nod, within normative ethical theory, to the social contract tradition: she mentions that John Rawls’ social contract theory is one proposal (100). Second, in Kant’s view the third formulation requires that we act as a citizen in the moral community. Korsgaard requires, instead, that we adopt the practical identity of a citizen in the moral community. She calls this practical identity “moral identity.” Whether this is a (serious) difference depends, of course, upon what it means to adopt a practical identity, which I will discuss in more detail later.}
What drives Korsgaard to this view? I think two motives. One, which I’ve already discussed, is that Korsgaard thinks that the universal law test and so Kant’s argument is subject to empty formalism objections. The reflective agent needs a practical point of view to make principled decisions, and there are immoral practical points of view that are formally consistent. The Categorical Imperative, by itself, simply doesn’t, and can’t, manage to force an agent to “veer right.” The agent who self-legislates a formally consistent practical identity has, in the very least, acted in a way that meets a “need” of rational agency, and there are formally consistent immoral principles of choice.

My suspicion is that Korsgaard has a second motive: she wants to “humanize” Kant’s theory. She is inclined to think that Kant’s theory of practical rationality isn’t sufficiently charitable. It judges people irrational who, at least in one important sense, are not. People who adopt and attempt to live up to normative self-conceptions, Korsgaard stresses, have responded to a “need” created by the “reflective structure of the mind.” In comparison with Kant’s universal law test, which Kant himself acknowledges to be abstract, Korsgaard’s “test of normativity” for the desire of the moment, she thinks, is concrete. And it is already how most of us think through our practical decisions: “We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves” (120). I take it that Korsgaard intends for the idea of a practical identity to appeal to her reader as intuitive: “You are . . . a woman or a man, an adherent of a certain religion, a member of an ethnic group, a member of a certain profession, someone’s lover or friend, and so on” (101). Each of these self-conceptions constitutes a role, with a set of practical guidelines “built in.” To Korsgaard’s credit, I imagine many people have had the kinds of thoughts she is gesturing at. What parent, for instance, hasn’t
had the thought, after a flash of anger directed at his children, “How can I call myself a parent?”

This a common and very human thought.

Does Korsgaard think, then, that the Categorical Imperative is a rule that is, in the very least, not often broken? It is conceivable to violate it, she points out. A pure wanton does, and a pure wanton is a logical possibility. (This leads Korsgaard to say that the Categorical Imperative isn’t empty, it could conceivably be violated and so is minimally demanding.) But my impression is that Korsgaard thinks that most of us have lived up to the demand stated by her Categorical Imperative. Achilles, for instance, did not merely adopt the Homeric ethic as a principle of choice, he invested great meaning into pursuing its ideals. He even preferred death to the staining of his personal honor. Achilles, then, had a practical identity. Korsgaard’s belief that most every human being has at least one practical identity leads to her judgment that her theory of practical reason is sufficiently charitable; it treats most people as minimally “rational and active” (238).

Korsgaard seems to press a bit further, for another point. Our common experience, she seems to think, is that we have “too many” personal obligations rather than “too few.” The reflective agent, pestered by the voice of reflection, caught in the midst of a bout of intransigence, might wonder whether any of the voices he hears deserve a voice; he is worried that there is nothing he has practical reason to do. But when it comes to you and to me, we usually have to deal with multiple legitimate voices. Speaking of practical identities, Korsgaard says “most of us have a jumble of such conceptions” (101). And you are likely to have a variety of self-conceptions, all at once: “You are a mother of some particular children, a citizen of a particular country, and adherent of a particular religion” (120). It seems to follow that what most of us experience, in so many of our choice situations, is not the practical analogue of Descartes’ feeling of disorientation and stultification that comes with having an empty self. We don’t experience a dearth of practical obligations, but a flood of them. We have more duties than we know how to fulfill, coming from multiple affiliations. What characterizes our day-to-day experience is a sense of obligation, of being pulled in various directions by the conflicting demands of our various self-conceptions. The voice of your practical identity speaks with authority within your practical deliberation. Though I think Korsgaard is exaggerating a bit when she says that “every obligation is a threat of a loss of identity” (101-102); there is a truth in what she says. It’s not unusual for a person to experience a conflict between two roles that she thinks important to herself and subsequently wonder whether she can be fully committed to both: Can I be both a father and a serious lawyer?
This is supposed to help Korsgaard’s argument. Recall, Korsgaard does not accept Kant’s “fact of reason,” which claims that people who misbehave morally are “without excuse” (405), for “reason presents” the Moral Law as “so deserving of respect.” Kant recognizes that a person might see this truth only through a glass darkly, because of the dear self and its deceptions or an inadequate moral education. But we—all of us—are sufficiently aware of the validity of the Moral Law to warrant the claim that we are without excuse if we misbehave. In misbehaving we are failing to treat the dictates of conscience as the genuine practical obligations that we know, in our mind of minds, they are. Korsgaard, however, accepts Bernard Williams’s objection (which was mentioned in chapter 4) that whether a person has a conscience—construed as a sense of the validity of the Kantian Moral Law—is culturally determined. ‘Did Achilles?’ we asked. Korsgaard is willing to say ‘no’. Consequently, her argument does not appeal to fact of reason to get an agent who finds in herself a dispute between conscience and a heart’s desire to veer right. The only ‘must’ the reflective agent hears when she is confronted by clamorous desires or demanding preachers is the ‘must’ “If I don’t have a practical point of view, I don’t have any way to ‘go forward’ or ‘be active and rational’.”

With respect to Kant’s fact of reason, Korsgaard’s argument does have an analogue. To leverage the uncommitted to a moral life, Korsgaard appeals to “the fact of value.” “It is the most striking fact,” she begins her book, “that we have values” (1). By ‘we’, Korsgaard does mean all conscious, adult, normally functioning human beings. What we—all of us—share throughout the ages is that we all have ideas about how we ought to live. (Achilles, too.) Hence, her argument is that the only way to escape the demands of the Moral Law is to become a “practical nihilist,” the view that nothing in the world is worth doing and nothing is worth living for. Perhaps many people have never thought explicitly about the option of becoming a practical nihilist, but if
presented with the option to reject the existence of values, she thinks people would be strongly inclined to continue to eschew practical nihilism and keep some set of values. We’d even be willing, she thinks, to continue to have a set of values if we became aware that having them has far reaching practical implications. One of the implications of having values, Korsgaard thinks, is that if you have any values whatsoever, if you make any choices whatsoever, you are rationally committed to moral rules. This is roughly what Korsgaard means when she says that “any reflective agent can be led to acknowledge that she has moral obligations” (125).

The “fact of value” and its “mystery” are accessible to anyone (2). An agent raised in parts of the world not much influenced by a Socrates or by enlightenment morality might well be unaware of the Moral Law and its validity. In Korsgaard’s view, the case for the legitimacy of moral rules is buried within difficult chains of reasoning, but—Korsgaard claims—it is there: “reflection doesn’t leave things just where they were” (257), and if “the light of reflection is on” (257), it will make an agent see that moral rules are valid for him. In short, even if Korsgaard intends to be more sensitive (than she takes Kant to be) to cultural factors in the development of personal values, and she is willing to describe a wide range of responses to the normative problem as rational—to a first approximation, I suppose.315

But she is also strikingly confident in the power of reflection to lead people to moral commitment. In this way, a kind of Kantian optimism about reflection and the authority of the Moral Law pervades her argument regardless of her disagreements with Kant himself. The Enlightenment project advanced by Kant is not merely an interesting cultural artifact; it has an important seed of truth, with welcome implications for morality. Not to mention, an adequate

---

315 Kant, of course, is also able to say that an agent that takes the means to his desired end is practically rational, “to a first approximation,” even if he doesn’t subject the desire’s maxim to the universal law test. The agent is means-end rational.
account of the rational impotence of immoral inclinations and normative self-conceptions.

Reason’s final remark is that they don’t really generate practical reasons, all things considered; a fully rational agent, lucidly aware of all the relevant information, would not treat immoral inclination as the source of practical reasons or an immoral practical identity as a valid law.

Heteronomy and Self-Legislation

Here we come to a question that I don’t know how to answer. Korsgaard takes herself to be a procedural realist: there is a proper way to self-legislate a principle of choice, or a practical identity. Does a principle of choice have to be anything more than internally consistent and determinate for it to become a law for an agent? It seems so.

Korsgaard does not reject Kant’s heteronomy thesis. She accepts the claim that the will has to be “entirely self-determining.” Her commitment to Kant’s heteronomy thesis means that she has to say that the will, the legislating self, is free only if it is not “conditioned” or “determined” by anything external to the will. So, presumably, she has to say that Achilles’ warrior ethic is properly normative for him only if his culture “influenced” but did not “condition” his will. That is to say, a practical identity is normative only if it is adopted in the right way, namely, not at the behest of a desire or anything else alien to the will itself.

Is it plausible to think that most agents are merely “influenced” but not “determined” by what Kant and Korsgaard count as forces alien to the will itself? One possibility is that a highly reflective agent who has confronted the problem of reflective agency recognizes that he needs some practical identity or another and simply opts for the one that his society has imparted to him, not because his will is determined by his upbringing, but because he recognizes that his contingently acquired values are is as good as any other. But isn’t it dubious to think that
Achilles went through this chain of reasoning or stepped back from his warrior ethic to seek an unconditioned answer to the normative question? Korsgaard needs to indicate what other type of behavior, short of this radically and highly self-conscious choice, is sufficient for proper self-legislation, consistent with autonomy. I don’t see where she answers this question.

If the procedure requires a high degree of awareness and self-conscious choice, presumably fewer principles that are treated as the source of practical reasons are legitimate sources. If it does not require some degree of self-conscious choice, how is this consistent with the heteronomy thesis? Not only that, it seems the choices themselves have to be made on the right grounds. I do think that Korsgaard’s idea that we are “involved” in the “construction” of our contingent self-conceptions is plausible, if read in the following way. Achilles, for instance, is clearly conscious of the costs of his practical identity; the preservation of his honor will cost him his life, and he walks to his death. His persistence suggests that he endorses his practical identity: he has chosen to follow it instead of his desire to live. He has, by an act of will, endorsed his practical identity. Our choices often play a role in either cementing our self-conceptions or weakening them. But it doesn’t follow from this that Achilles’ act of will is consistent with the heteronomy thesis. He might, with respect to Kant’s conception of freedom, be “nothing more than an arena” in which society’s values have taken hold. He is then living up to them, without being reflective about them. Or he might care deeply about being a warrior.  

316 I think there are a set of content-independent considerations that make it rationally advisable to have a practical identity. Your life would be very boring if you didn’t have projects to pursue, and it would feel less meaningful if you didn’t have things to care about. These claims are central to Frankfurt’s account that human beings have a very strong practical reason to adopt personal projects and “care about something.” But Frankfurt grounds these commands in contingent motivational states: we hate boredom, it’s a form of suffering, and we don’t want to suffer.

316
A Tension

The two stages in Korsgaard’s argument pull in different directions, and it will be an important task for Korsgaard to reveal how they fit together. The first stage of Korsgaard’s argument attempts to pay respect to what is contingent in human life (241). Each of us is born into a particular community, into a particular family, perhaps in a particular ethnic group, and we “stumble into” our deepest values, often deeply influenced by cultural and personal factors (241): “Our contingent [values] are, to some extent, given to us—by our cultures, by our societies and their role structures, by the accidents of birth, and by our natural abilities” (239).

In Korsgaard’s view, these contingent social and personal influences are commonly psychologically powerful, and they are generally a part of the true explanation for how we come to have the values we do. These influences play a crucial role in shaping a person’s identity, “who he is” and “who he takes himself to be.” But Korsgaard also thinks that the contingent values an agent “stumbles into” are commonly normative for him. Presumably, given her commitment to the heteronomy thesis, she thinks that they are commonly properly endorsed by the agent’s will; if so, they are the source of practical reasons, even obligations (241). For Korsgaard thinks that your values are often the consequence of a mixture of the social influence and personal choice; for she says, “but it is also clear that we enter into their construction” (239). Achilles was deeply influenced by his culture. But he really did have a practical obligation to protect his honor above all else, even to kill for it if his warrior ethic so prescribed. There is a charity here that calls to mind, well, the inclusiveness of the subjectivist tradition.

The second stage of Korsgaard’s argument presses us to recognize what is universal. We all share one identity in common: we are beings with the capacity for reflection, and who consequently “need” the kind of values that society and family and the accidents of our own
peculiar personal history provide. Our capacity for reflection forces us to have values: without them, “you will lose your grip on yourself as having any reason to do one thing rather than another” (121). Korsgaard thinks that reflection upon this universal feature of humanity—this “need,” and its source in our capacity for reflection—ultimately leads us, if we are lucid and reasonable, to a proper respect for the human community as a whole and each of the members within it.

Achilles, Korsgaard has indicated, has a genuine obligation to do what his practical laws tell him. He might not have recognized that the Moral Law is valid for him. But Korsgaard thinks that he ought to have. He ought to have recognized that he had another and “deeper” identity, “human identity,” in which his contingent warrior identity is “grounded.” As Korsgaard would put it, Achilles’ practical obligation to kill for the sake of his own honor derives from a practical law that is not “reflectively stable” (128). If he were to think lucidly, his endorsement of this practical law would not “survive reflection”: “reflection does not leave everything where it was.” In this way, what he is guilty of is being “insufficiently reflective” (250): “Guided by reflection, we may be led to see that our tendency to treat our contingent practical identities as the source of reasons implies that we set a value on our own humanity and humanity in general” (250).

Notice that Korsgaard shares Kant’s thesis that there is, in the human community, important diversity, with strong grounds for solidarity. (This is a thought I find profoundly important for moral and political thinking.) There is proper diversity—at least some of my

---

317 Korsgaard says, “it is important to remember that no argument can preserve any form of relativism without at another level eradicating it” (118). Is this true? Take ethical cultural relativism, the normative ethical theory that claims that what makes a morally right action, A, morally right for a person, P, is that A is prescribed by the culture in which P lives. If ethical cultural relativism is true, it is true for everyone. But this doesn’t “eradicate” the relativism. To say so is to conflate the truth of a claim and its content: a true claim’s truth is “universal”; it’s only the content of the standards of moral rightness that is relative.
projects and goals are likely different from yours, and from Achilles’—but we all share a human identity, the capacity for reflection, the exercise of which Korsgaard regards as the source of human dignity. Korsgaard wants to treat a wider range of maxims and acts of self-legislation as practical-reason-giving than Kant. The Categorical Imperative can’t, by force of reason, bring us together into a kingdom of ends. It takes another principle, and another argument, to do that. But she shares Kant’s view that conscience always deserves a voice in practical deliberation.

My attribution of a desire to be charitable helps to explain why Korsgaard chooses the relativistic language. Her Categorical Imperative mimics one way in which moral cultural relativism is more permissive than moral universalism. Moral cultural relativism gives a significant role to cultural factors in the generation of moral rules. Korsgaard’s theory of practical reason seems to give a similarly significant role to the contingent values of, among other things, an agent’s culture in the generation of the agent’s personal obligations. An agent’s cultural heritage is often, in Korsgaard’s theory, a genuine source of practical reasons even if the values of his culture do not cohere with the values of Kant’s enlightenment morality.

Given her commitment to Kant’s heteronomy thesis, though, Korsgaard will have to explain what type of influence is consistent with personal autonomy. Korsgaard rejects reflective endorsement from the perspective of your heart’s desire, or from the perspective of your contingent values themselves, as a legitimate way to become personally autonomous. If Achilles’ endorsement of the warrior ethic is nothing more than an endorsement from the set of values instilled into him by society, then it would seem Achilles isn’t, according to Kantian standards,

---

318 This seems as though it is relativism because there will be, according to Korsgaard’s theory, cases in which two persons, or two groups of people, make deeply conflicting claims about what there is practical reason to do, and both will be correct. That’s no trouble for subjectivism about practical reasons, but it isn’t typically Kantian.
personally autonomous. (As a subjectivist, I think he could be, if these values are ones he cares about.) He is, to use Korsgaard’s phrase, nothing more than “an arena in which” the values of society are operating. His will is not “entirely self-determining.”

The Unity of Practical Reason

The sense in which the two stages of Korsgaard’s argument seem to pull apart is that the first stage attributes to the will a level of freedom and practical authority that the second stage attempts to constrain. And so, the Categorical Imperative endorses the very principles of choice that the transcendental principle rejects. But both the Categorical Imperative and the transcendental principle are rules of rational willing. The practical reasons that derive from acts of self-legislation that comply with the Categorical Imperative are not merely *prima facie* practical reasons; they are “genuine” practical reasons (257). The Categorical Imperative permits ethical egoism and the Homeric warrior ethic; if properly self-legislated, each becomes a genuine practical law for an agent, a source of genuine practical obligations for him. The will, recall, “is practical reason,” and so practical reason endorses these principles of choice. But the transcendental reasoning of the second stage is supposed to condemn ethical egoism and the Homeric ethic—or, at least, the aspects of the Homeric ethic that do not accord with the Moral Law. The test of reflection in the second stage is supposed to lead to the judgment that lucidity reveals that ethical egoism and important aspects of the Homeric ethic aren’t really legitimate sources of practical reasons after all. Otherwise put, the legislating aspect of the will is free, practical reason says, to incorporate any maxim into the personal will, but there are some maxims, practical reason says, that it really should not. Does practical reason itself speak with two voices?
The question is how Korsgaard intends to fit these judgments together in a single, coherent theory of practical rationality: does practical reason speak with a single voice? How so, when it claims, of one and the same principle, for one and the same person, that it is both creative of genuine practical reasons, and not?\footnote{Korsgaard’s theory is especially difficult to make sense of, since she seems to deny that the practical principle of non-contradiction is absolute. She claims that what “makes life hard” is that genuine personal obligations conflict, and yet she wants to press the ethical egoist to give up his egoism on the grounds that it is inconsistent with moral identity, which is rationally inescapable. Korsgaard will have to indicate why the ethical egoist does have to render his will consistent by giving up his egoism.} How does the will’s authority, which cannot be questioned, cohere with the universality of the Moral Law? Korsgaard will need to explain how these claims fit together.

I am willing to make a proposal, but it’s not friendly. I do think it is coherent to ask, of any basic desire, whether it is rationally potent, and I do think that there are instances in which the answer is ‘no’. I also agree that a normative self-conception is able to speak against certain basic desires. The so-called “relativism” seems just right. I also happen to think that the conception of personal autonomy Kant and Korsgaard advocate is closely connected to a very important ideal, an ideal I personally cherish. There is a form of existential maturity that requires stepping back from and critically reflecting on our mental states. These thoughts can be brought together into a consistent, plausible theory. But I don’t think we should make this ideal of personal autonomy a part of the very logic of action, and that’s what the heteronomy thesis presses Kant and Korsgaard to do.

The Argument for the Universal Validity of Moral Identity

In the second stage of her argument, Korsgaard attempts to argue \textit{from} the Categorical Imperative \textit{to} the validity of the Moral Law, that is, to the universal authority thesis. Korsgaard
claims that whatever maxims a particular agent happens to have (properly) self-legislated, his practical commitment to them rationally commits him to “moral identity,” the practical identity of a citizen in a republic of ends. As mentioned, Korsgaard needs to supply at this stage in her argument a test of practical reason—distinct from the universal law test—that takes properly self-legislated maxims as proper input. This further test will itself need to be a formal test, given that Korsgaard claims, as Kant does, that the rules of practical reason are entirely formal and without substance and yet lead, by some route or another, to the validity of the Moral Law. This test of practical reason has to include, though, more than rules for volitional consistency. Otherwise her argument would suffer from an empty formalism objection, too. There are formally consistent immoral and amoral self-conceptions. To this end, Korsgaard appeals to a transcendental principle.

This is one point at which I find the argument sketchy. I can lay out the sketch, but I’m not sure how Korsgaard would fill in the argument. To paraphrase, Korsgaard thinks that self-legislated, and acting from, a maxim implicates you in “valuing.” To self-legislate a principle of choice and to treat its practical advice as the source of practical advice for you is, she thinks, a type of valuing: you are treating it as normative for you. Since this type of valuing is only possible because we have the capacity for reflection, and the capacity for reflection is what makes us human beings (in her technical sense), then we are, she reasons, rationally committed to valuing our own humanity. Since I have to value my humanity, and your humanity is no different than mine, a principle of consistency in judgment implies that I have to value yours, too. If I value each and every person’s humanity, this is tantamount to being obedient to the Moral Law. That’s how the argument works, in broad outline.
One strength of Korsgaard’s argument, in comparison with Kant’s, is that there is no grounds—at least so far as I can see—for suspicion that Korsgaard’s argument appeals to thick assumptions from the natural law tradition or to controversial doctrines from a religious tradition. (If anything, the opposite: Korsgaard calls the moral community a “republic” instead of a “kingdom” of ends because she does not think this community has a king, God.) She doesn’t posit, for instance, the theological claim that human beings are meant to be—given that we are essentially creatures, not God—less than fully-self sufficient. Since she is not attempting to derive substantive results from a universal law test, there is less temptation to sneak details in. Consequently, Korsgaard seems to escape the fourth set of objections to Kant.

That said, Korsgaard does have her own striking commitments. There is the transcendental principle, which goes something like this: If you value X, then you are rationally committed to valuing whatever makes X possible. Is this principle true? I find the language of “valuing” vague. What type of valuing does my treating a practical identity as normative, or do my acts of self-legislation, commit me to? It is clear that Korsgaard wants to argue that I must value other members of humanity, as a thing of objective, absolute, and intrinsic worth: as an “end in itself” (125). This is what the Moral Law, in its second formulation, essentially demands. Do my practical identity and acts of will commit me to valuing myself as an object of objective, absolute, and intrinsic worth? It is true that without my capacity for reflection, I would not be able to identify myself with a practical standard such as a normative self-conception, and I would not be able to engage in acts of will; my capacity for reflection is a “necessary ground” for these activities. Does the logic of agency entail that I treat every “necessary ground” of something I “value” as intrinsically valuable? That seems unlikely. The person who values his own
temperate behavior needn’t intrinsically value the temptations that make his exercise of temperance possible, must he?

Korsgaard also attributes some surprising things to the reflecting self, who is not, in her theory, a “thin, wispy” ego, with little more than the ability to step back and gaze. She thinks the reflecting self itself has not only the ability but the practical authority to compel an agent to seek answers to the normative question, and she thinks the “acting self” is under an obligation to obey.

What if the Mafioso puts his mind to thinking about how to live up to his identity, but never really thinks about whether he ought to have the practical identity that he employs as a practical guide? Korsgaard claims that he cannot properly ignore the normative question, or any of the questions that reflection asks:

the activity of reflection has rules of its own . . . And one of them, perhaps the most essential, is the rule that we should never stop reflecting until we have reached a satisfactory answer, one that admits of no further questioning” (257-258).

The reflecting self is not as empty, it turns out, as Korsgaard has so far suggested. The reflecting self has the “right’ to command you to ask for practical justification of your action-guiding self-conceptions.

Myself, I wonder where this “right” comes from. No doubt, reflection does, as a matter of fact, raise questions, but why think it has “right” to do so? Is it simply built into the fabric of things, similar to a realist standard, that the reflecting self has the practical authority to obligate you and I to follow out chains of questions “until we have reached a satisfactory answer”? I’m not sure. Korsgaard doesn’t indicate where this practical authority comes from. Her language at times suggests that it would be profoundly inauthentic not to listen to the voice of reflection, since you cannot deny that one thing that you are is a creature with the capacity for reflection. This is an incorrigible belief; does that make it deep?
Questions and Objections

It seems to me that Korsgaard is not able to do any better than Kant at meeting this challenge. And in the end, in an attempt to make her argument go through, Korsgaard is forced to make claims about the logic of rational agency as extravagant and uncompelling as Kant’s commitment to weak and strong universalism. Korsgaard’s appeal to a transcendental principle is subject to serious problems. It’s not clear to me, for instance, how the obligation to reflect arises. Is it simply built into the fabric of things? Is it a rule constitutive of reflection?

Korsgaard admits that her argument is sketchy. And I think there are several points at which it is difficult to tell how she means for it to work. What follows is several important questions that Korsgaard needs to answer to make her argument more clear and to overcome several of my objections.

First, how does the will, the legislating self, make its initial choice between principles of choice? If it is to behave as a will, it has to be “entirely self-determining,” and so it can’t be “conditioned” by an alien cause. A desire or an internalized norm, inculcated by enculturation, is able to operate as an “influence” but not a “determinant” of the will. The legislating self would not be free if it were simply enacting laws on behalf of a desire of the moment, or a heart’s desire, or in deference to the demands of another person or society. And Korsgaard considers the will, in its role of self-legislation, to be free of moral constraints. So moral constraints are able to operate as practical laws only if they have been properly endorsed by the will. And I’m asking right now how the “first” act of self-legislation is made. The will, it seems, has a lot of principles of choice to choose between. There are, as discussed in the chapter about Kant, a very large number of formally consistent principles of choice. But how is it to choose?
Second, there is also the question why the will has the authority to make laws. Where does its (absolute) practical authority derive from? I can imagine instances in which an act of will is useful, or even necessary. Perhaps I would have trouble sticking to a personal project that I really care about if I do not endorse the project by an act of will. I don’t want to feel like a quitter, and so I resolve to stick to the program. In this case, I am tapping into my hope not to feel guilty for slacking, and using it to keep myself to the program. But isn’t this a case in which the will is serving the interest of the heart’s desire?

Third, why doesn’t Korsgaard make a more economical argument for the Moral Law? Korsgaard appeals to a transcendental principle, and she thinks she is able to argue that the “value of humanity itself is implicit in every human choice” (122). I don’t know how, exactly, Korsgaard prefers to articulate the relevant transcendental principle. But let’s say that the principle is to be stated in the following way: ‘you are rationally committed to valuing whatever makes possible the things you do value’. If this is roughly how she would express the transcendental principle, why can’t she argue in the following way? She might begin from the claim that every action is done from a principle of choice, and acting from a principle of choice rationally commits you to value; it implicates you in valuing—to thinking the principle is the source of a practical reason. Then she could point out that your capacity for reflection is a precondition for acting or choosing or valuing, and so—by way of the transcendental principle, you have to value your capacity for reflection. The transcendental principle is expressed, as it often is, generically in terms of valuing: “an agent has to value whatever makes possible the things he values.”

This principle has provoked some Kantians to argue that any human choice implicates the agent in valuing. Permit me to speak somewhat loosely, the argument tends to go something like
this. When you pursue an end, this implies that you value it. And since having the capacity for choice is a precondition for pursuing any valued end, an agent has to value his capacity for choice. Since the capacity for choice is a defining characteristic of humanity, an agent, in any choice, is implicated in valuing his humanity.

Korsgaard takes us, instead, through a discussion of action-guiding self-conceptions. Is commitment to a practical identity present in every human choice? (The “under which” contained in her gloss of ‘practical identity’ calls to mind Kant’s claim that each action is done “under the idea of freedom.”) Does Korsgaard mean to say that every action is done “under the idea” I am a such and such, where ‘such and such’ is to be replaced by the name of a relevant practical identity?
CHAPTER 7
CONCLUSION

According to my interpretation of Harry Frankfurt’s work, Frankfurt is a refined subjectivist about practical reasons who accepts the rational impotence thesis and needs to explain how basic desires are rendered impotent, and in a way that is consistent with subjectivism. I think my proposal has. It reveals, in an intuitive and theoretically economical and modest way, sensitive to the details of personal ideals that some agents cherish, how the voice of an agent’s will is able to disenfranchise desires.

I do agree with several aspects of Korsgaard’s theory. Her appeal to the concept of a normative self-conception, for one. And I don’t have a problem with Korsgaard’s “relativism” about practical reasons. As a subjectivist, I think practical reasons are “agent-relative,” and so I think subjectivism is able to be charitable in the way Korsgaard wants to be. A subjectivist is able to say that Achilles acted from “genuine” practical reasons. On the assumption that Achilles cared about the Homeric ethic in the proper way, it is a valid practical standard for him, whatever influence society has or hasn’t had upon him. Subjectivism does not accept Kant’s heteronomy thesis, and so it says that a person is able to be self-governing by reflecting upon the demands of his heart’s desire and pursuing its satisfaction. It is able to do this without generating the false “crisis” of authority the reflective agent, in Korsgaard’s theory, faces; and without giving, implausibly, absolute authority to the will to simply pick out principles of choice. Whereas
practical reason commands, in Korsgaard’s theory, “adopt a law, any law”; in mine, practical reason tells us to shape our practical standards in accordance with what we care about, to find the most effective ways to pursue what is dear to us.

Though Korsgaard’s remarks about the concept of a practical identity are terse and sketchy, I think she is gesturing at an important and useful concept. Our best model of practical deliberation and rational agency ought to include such an idea. I disagree with the claim, which Korsgaard seems to make, that a practical identity is necessary for the exercise of rational agency. But I think some practical deliberation is done from the standpoint of a personal ideal; people do sometimes care about what kind of person to be and aspire to live up to the standard such an ideal presents. In my view, such ideals often contain norms of rational impotence. I don’t think every practical identity does, but some do. A role such as being a Christian or a psychiatrist or a pacifist is the source of norms not only about how (not) to behave but also which inclinations to treat as rational (im)potent. There are some things, say, a pacifist can’t properly do—go to war, for instance—and there are some desires a pacifist can’t treat as rationally potent—a vengeful impulse, for instance.

As I see it, agents are able, and often do, act merely to satisfy a simple inclination, such as a desire for ice cream, which is not grounded in a self-conception. My claim is merely that some actions derive from more complex types of practical deliberation. Our best models of practical deliberation and rational agency ought to make room for whatever shape the exercise of practical reasoning might take. Even though simpler models of agency capture some types of behavior, our full-fledged model ought to be able to capture any. So, there is important work for such a concept to do. On one reading of Frankfurt’s example, he, as a father, feels the compelling force of the norms of rational impotence contained in his own personal ideal.
BIBLIOGRAPHY


