ABSTRACT

Neo-Kantian Wickedness:
Constructivist and Realist Responses to Moral Skepticism

Heidi Chamberlin Giannini, Ph.D.

Committee Chairperson: Robert C. Roberts, Ph.D.

Neo-Kantian constructivism aspires to respond to moral skepticism by compelling agents to act morally on pain of irrationality. According to Christine Korsgaard, a leading proponent of constructivism, we construct all reasons for action by following correct deliberative procedures. But if we follow these procedures we will find that we only have reasons to act in morally permissible ways. Thus, we can show the skeptic that he is rationally constrained to act morally. Unfortunately, as I argue in my first chapter, this strong response to moral skepticism renders deliberate immoral action unintelligible. This result is problematic since we often do interpret ourselves and others as deliberately choosing to do wrong. I further suggest that this problem follows from central commitments of Korsgaard's constructivism, so that any adequate account of immoral action must abandon constructivist metaethics in favor of moral realism, a suggestion reinforced by the argument of my second chapter. There, I call attention to Kant's solution to a similar problem in his own account of morality. I argue that Korsgaard's constructivist commitments prevent her from embracing Kant's solution. I proceed in my third chapter to argue that there is a further tension between Korsgaard's response to

moral skepticism and her work in non-ideal theory. In particular, Korsgaard maintains that, when confronted with injustice, the virtuous person may have reason to do what is wrong in the name of morality. She thus relies on the assumption that one can *deliberately* do wrong. I argue that this assumption undermines the response to skepticism that motivated Korsgaard's constructivism in the first place. But despite the problems with constructivism, we may worry that moral realism fails to offer an adequate response to moral skepticism. Indeed, Korsgaard rejects realism in part because she believes that realists simply refuse to respond to moral skepticism. I thus conclude by arguing that moral realists can offer adequate responses to moral skepticism. In fact, I believe Korsgaard's response is no more effective than those suggested by some moral realists.

Neo-Kantian Wickedness:
Constructivist and Realist Responses to Moral Skepticism

by

Heidi Chamberlin Giannini, B.A., M.A.

A Dissertation

Approved by the Department of Philosophy

_____

Michael D. Beaty, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Philosophy

Approved by the Dissertation Committee

_____

Robert C. Roberts, Ph.D., Chairperson

_____

Kyla S. Ebels Duggan, Ph.D.

_____

Thomas S. Hibbs, Ph.D.

_____

Michael D. Beaty, Ph.D.

_____

Darin H. Davis, Ph.D.

_____

Jonathan Tran, Ph.D.

Accepted by the Graduate School

May 2013

_____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

Copyright © 2013 by Heidi Chamberlin Giannini

All rights reserved

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

DM	*On Evil*
G.	*Groundwork of the Metaphysics of Morals*
MS	*The Metaphysics of Morals*
KpV	*Critique of Practical Reason*

PREFACE:

Wickedness and Moral Skepticism

My primary philosophical interests concern the concept of moral transformation: what is it? How is it possible? What can we do to bring it about? Given my interest in moral transformation, I am committed to taking seriously the idea that we do not start out in life as fully-formed, virtuous agents. More than that, I want to take seriously the idea that we sometimes *deliberately* choose to do things that are wrong. Perhaps this desire is partly responsible for my ongoing fascination with the thought of Immanuel Kant, who was notoriously suspicious of human nature. Unfortunately, as we will find over the course of this dissertation, the need to make room for the possibility of deliberate, wrong action is in tension with another goal embraced by some Kantian moral philosophers: the aspiration to respond to moral skepticism.

In *Religion within the Boundaries of Mere Reason*, Kant distinguishes between three types of moral failure: frailty, impurity, and depravity – sometimes translated as wickedness.[1] According to Kant, wickedness results when the agent chooses to privilege self over adherence to the moral law. Thus my use of the word "wickedness" in the title refers to this propensity to deliberately choose to do the wrong thing. It is important to note that deliberately choosing to do the wrong thing need not be the same thing as choosing to do the wrong thing *because* it is wrong. Kant thinks the latter – what he calls

[1] See, for example, Immanuel Kant, *Religion within the Limits of Reason Alone*, trans. Theodore M. Green and Hoyt H. Hudson (New York: Harper & Row, 1960), 25.

the "diabolical will" – is not possible for human beings. I have no interest in arguing that it is possible or impossible for human beings to choose to do the wrong thing because it is the wrong thing. Rather, I only want to argue that ethical theory must make room for human beings to be able to choose to do the wrong thing for *some* reason.

Ironically, I believe some of Kant's contemporary proponents, the neo-Kantian constructivists, are guilty of rendering deliberate immoral action impossible. Neo-Kantian constructivists are motivated largely by a desire to provide a response to moral skepticism – that is, skepticism about whether there is most or overriding reason to be moral.[2] I will show that in their attempt to demonstrate that there is most reason to be moral, the constructivists maintain that agents have *no* reason to do immoral things. In this way, they render deliberate immoral action impossible.

It is worth observing that "constructivism" can refer to a broad class of positions within moral philosophy and that I am concerned only with a specific subset of those positions. Constructivism, most fundamentally, is simply the view that some feature of morality or normativity (certain types of obligations, principles, values, reasons, etc.) is constructed by human agents. Thus, different versions of constructivism can be more or less radical. For example, T. M. Scanlon is sometimes described as a constructivist because he believes moral obligations are constructed out of more fundamental reasons,[3] but his constructivism is not my target in this dissertation. Indeed, I will treat him as an

[2] I will use "most reason" and "overriding reason" interchangeably. I don't mean to take a position on whether reasons to act morally are weighed against other reasons or override other reasons.

[3] See Carla Bagnoli, "Constructivism in Metaethics," *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/constructivism-metaethics/ (accessed January 14, 2013).

ally, an opponent of the sort of constructivism I critique. For the purposes of my dissertation, "constructivism" will refer only to the subset of constructivist positions that, in the words of Christine Korsgaard, "go 'all the way down.'"[4] That is, I will target views that deny that there is any "unconstructed" normative truth. While Scanlon believes moral obligations are constructed, those obligations are constructed out of reasons that, though normative, are *not* constructed. Korsgaard, by contrast, proposes a view on which all normativity is constructed.

Several contemporary philosophers endorse constructivist views of this sort, but no one has taken on the project of articulating and defending this approach to ethics so comprehensively as Christine Korsgaard. Thus, I will rely on Korsgaard as my interlocutor. I will begin by showing how Korsgaard, in her attempt to respond to moral skepticism, rules out the possibility of deliberate, immoral action. I will also suggest that making sense of deliberate, immoral action requires us to embrace an alternative view of the nature of normativity: realism. Furthermore, I will illustrate how Korsgaard's attempts to make room for deliberate wrongdoing compromise her response to skepticism. However, one of Korsgaard's critiques of realism is that it fails to address moral skepticism. In my last chapter, then, I will argue that moral realists are capable of providing an adequate response to moral skepticism.

Responses to moral skepticism can be in tension with attempts to account for deliberate immoral action. However, I think it is possible to explain how people can sometimes choose to do wrong without compromising our ability to answer the moral skeptic. And I believe doing so requires us to embrace moral realism.

---

[4] Christine M. Korsgaard, "Realism and Constructivism in Twentieth-Century Moral Philosophy" in *The Constitution of Agency* (New York: Oxford UP, 2008), 302-326. See p. 324.

ACKNOWLEDGMENTS

I would like to thank the members of my committee for their hard work and helpful feedback, especially my director, Bob Roberts, and Kyla Ebels-Duggan, who exhibited tremendous commitment to the completion of this project. She expended more time and effort in assisting me than perhaps anyone else, going so far as to allow me to live in her home for a semester while I wrote. I would also like to thank the Baylor Philosophy Department for the support and resources I needed to finish. In particular, I'd like to mention my department chair, Mike Beaty, who allowed me to take advantage of several unique opportunities that enriched my education, and the members of my dissertation writing group, David Echelbarger, Nate Carson, and Nate Jackson, who commented on multiple drafts of my writing and held me accountable throughout the writing process. Special thanks are also due to the Northwestern Philosophy Department for taking me in for a semester while I worked on my dissertation.

Of course, I could not have finished my dissertation without the love and support of my friends and family. Three people in particular deserve mention. My father was a constant source of agitation, asking me every time he called if I was working on my dissertation. But that agitation was also a tremendous source of motivation. Thank you for the support, Dad. Janelle Aijian has been my closest friend all through graduate school. These last seven years would not have gone nearly so well without her sympathizing with me every step of the way. I would also like to thank my husband,

John. He came into my life at just the right moment. I don't know if I would have been able to finish without him.

Finally, as much as I resist sounding trite, I would like to thank God. He has been so good to me. His power is made perfect in my weakness. All honor and glory to Him.

CHAPTER ONE

The Problem

> Contemporary secular ethicists do not [pay attention to viciousness], as a rule, but some of what people deliberately do is very unethical, and blinding ourselves to what can be said on behalf of what people in fact deliberately do is blinding ourselves to what they are up to in action, the *nature* of their purposes. This is no minor limitation for a theory of practical reason.[1]
>
> --Candace Vogler

Introduction

In 1996, Christine Korsgaard sought an answer to what she called 'the normative question.' Having provided an account of the content of right action, she claims that the moral philosopher is still confronted with the question: "must I really do this? Why must I do it?"[2] The answer to this question, which Korsgaard attempts to supply, would both silence the moral skeptic who doubts whether he must act morally and provide insight into the nature of obligation. But Korsgaard's answer was promptly accused of describing "the experience or phenomenology of obligation, not its ground or authenticating source."[3] Korsgaard's account of normativity has developed significantly since then, but I think there is still some truth in this accusation. I begin by outlining her earlier work on the normative question and explaining the above objection to it. I will

---

[1] Candace Vogler, *Reasonably Vicious* (Cambridge, MA: Harvard UP, 2002), 37.

[2] Christine M. Korsgaard, *The Sources of Normativity* (New York: Cambridge UP, 1996), 16.

[3] G. A. Cohen, "Reason, humanity, and the moral law" in *The Sources of Normativity*, ed. Onora O'Neill (Cambridge UP, 1996), 167-188.

then proceed to discuss her more recent work and why I think a version of this objection remains unanswered.

Realism and Noncognitivism

At the heart of Korsgaard's work is a rejection of both realist and noncognitivist views of reasons for action. According to realist accounts, reasons correspond to facts about what it is good, bad, right, or wrong to do. These facts exist in the world independently of human reason. Noncognitivists, by contrast, think that reasons for action are located in the mental states of agents. Korsgaard objects to both views because she believes – and she thinks "most philosophers would agree" – that reasons for action have the following three properties: "(1) They are normative, that is, they make valid claims on those who have them. (2) They are motivating, that is, other things equal, the agents who have them will be inspired to act in accordance with them. And (3) they are motivating in virtue of their normativity; that is, people are inspired to do things by the normativity of the reasons they have for doing them."[4]

Korsgaard rejects the realist position because she believes that facts can be neither normative nor motivating. The noncognitivist agrees that facts cannot motivate action and that is why he insists that reasons consist in an appropriate mental state, e.g., the desire to do X. Unfortunately, if the existence of reasons depends only on the mental states of a particular agent, it is not clear how such reasons could be normative. There is nothing guiding about having the mental state that you have. That is, you cannot be right

[4] Christine M. Korsgaard, "Acting for a Reason" in *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (New York: Oxford UP, 2008), 207-229. Actually, she says, "Most philosophers would agree that practical reasons have at least *some* of the following properties." Korsgaard, "Acting for a Reason," 208, italics my own.

or wrong in having the mental state that you have, so having a certain mental state offers no guidance as to how you *should* act. A related objection attacks this view for making morality contingent: one has reason to conform to the demands of morality only if one has the requisite antecedent beliefs and desires. This view does not have the resources needed to answer the moral skeptic as Korsgaard wants.[5]

The defender of a realist account of reasons will contend that these mental states are only the appropriate responses to reasons: they are not the reasons themselves. Reasons motivate by provoking these responses. Furthermore, to have the appropriate response to reasons is just what it is to be rational. But this claim makes Korsgaard uneasy. After all, animals can respond 'appropriately' to facts, as when a lioness defends her cubs from danger. While Korsgaard has no problem attributing agency to non-human animals,[6] she does think there is a difference between rational human action and the behavior of the lioness, namely, "the human being is aware of the reason *as a* reason."[7] She explains, "[Rational action] is a matter of being motivated by the awareness or belief

[5] Korsgaard tells a story about two such competing philosophical camps in several places, presenting her own position as a kind of middle way. Cf. *The Sources of Normativity*, 21-89; "Kant's analysis of obligation: The argument of *Groundwork I*" in *Creating the Kingdom of Ends* (New York: Cambridge UP, 1996), 43-76; "Aristotle and Kant on the source of value" in *Creating the Kingdom of Ends* (New York, Cambridge UP, 1996), 225-248.

[6] I will later criticize Korsgaard for arguing that morality is a constitutive standard of agency and, as a result, that immoral behavior cannot be deliberately chosen by an agent. But it might be thought that she could still attribute some agency to wrongdoers since she clearly believes there are different kinds of agency, such as the agency of human beings and the agency of non-rational animals. I do not think this is a satisfactory response for two reasons. First, it would still mean that immoral behavior is at best the product of non-rational animal agency. But we don't regard all wrongdoers as 'animal-like.' Second, I think she would agree that the principles of human agency are simply different from the principles of non-rational animal agency. Even wrongdoers are *rational* animals and their agency must be grounded in the principles proper to the agency of rational beings. See Christine M. Korsgaard, "Fellow Creatures: Kantian Ethics and Our Duties to Animals," *Tanner Lectures on Human Values* 24 (February 2004): 77-110, http://dash.harvard.edu/bitstream/handle/1/3198692/korsgaard_FellowCreatures.pdf?sequence=2 (accessed January 15, 2013).

[7] Korsgaard, "Acting for a Reason," 214, emphasis hers.

that these facts *constitute* good-making properties of the action. To act rationally is to act from the belief that what you are doing is in some way good."[8]

An interesting conclusion follows from this. Korsgaard claims that "to say that the facts constitute good-making properties … is just to say that they provide the agent with what the agent regards as appropriate grounds for motivation."[9] But this in turn means that rational action is essentially being "motivated by the idea that your motives are good."[10] To clarify how this works, Korsgaard introduces a distinctive conception of action she believes is shared in common by both Aristotle and Kant: actions are "acts-for-the-sake-of-ends."[11] Killing oneself in order to save others, Korsgaard says, is different from the action of killing oneself in order to avoid hardships. The significance of this definition of an action is that "the *aim* is included in the description of the action, and … it is the action as a whole, *including the aim*, which the agent chooses."[12] The aim, then, is not the *reason* for the action; it is a *part* of the action. When you choose a course of action, you choose it for its own sake. Korsgaard concludes, "The person who acts for a reason, like God in the act of creation, *declares* that what he does is good."[13]

It is easy, at this point, to wonder whether such reasons are normative in any meaningful sense. Can't someone simply declare *any* action good? Korsgaard thinks

---

[8] Ibid., 214.

[9] Ibid.

[10] Ibid.

[11] Ibid., 219. Much of the following material drawn from this article is also present in the first chapter of Christine M. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (New York: Oxford UP, 2009).

[12] Ibid., 218.

[13] Ibid., 229, italics hers.

not. She believes that when we ask for a reason for an action, we are asking the agent to explain what makes a certain act worthwhile. She says,

> [T]he practice of answering the motivational question 'why?' by citing the agent's purpose does not really suggest that what we choose are acts, and our reasons are provided by our purposes. It is just that the purpose is often … the missing piece of the agent's maxim, the piece we need to have in place before we can see why the agent thought that this action as a whole was a thing worth doing.[14]

If someone who lives in Indianapolis goes to Chicago, and we ask why, the response "to buy a box of paperclips" is not an acceptable reply because it would not be *worthwhile* for a person to travel that distance to perform such a menial task.[15] But what it is worthwhile for a person to do is tied to *who* a person is: I can explain that I am reading Elizabeth Anscombe's *Intention* because it is of central importance to work in action theory, but if my younger brother – a command post controller in the Air Force with no philosophical hobbies – were to offer the same explanation for reading *Intention*, we would need some more information to understand why it is worthwhile for *him*. Who a person is, including the various projects he undertakes, determines his reasons. Not only can identity place limits on what counts as a reason, but Korsgaard believes it can even dictate what things a person is *required* to do.

Korsgaard's Early Work

When G. A. Cohen first accused Korsgaard of describing the experience of obligation rather than its source, she claimed to locate the source of obligation in the agent's self-conception or identity. She explains, "Reflective distance from our impulses makes it both possible and necessary to decide which ones we will act on: it forces us to

[14] Ibid., 222.

[15] Ibid., 221.

act for reasons. At the same time, and relatedly, it forces us to have a conception of our own identity, a conception which identifies us with the source of those reasons. In this way it makes us laws to ourselves."[16] Cohen believes this maneuver explains nothing more than the *feeling* of being obligated because we can imagine persons who feel obligated to act in certain ways in light of their conceptions of themselves but whose actions we would not want to admit are in fact obligatory. For example, he asks us to imagine 'an idealized Mafioso:'

> This Mafioso does not believe in doing unto others as you would have them do unto you: in relieving suffering just because it is suffering, in keeping promises because they are promises, in telling the truth because it is the truth, and so on. Instead, he lives by a code of strength and honour that matters as much to him as some of the principles I said he disbelieves in matter to most of us. And when he has to do some hideous thing that goes against his inclinations, he steels himself and we can say of him as much as of us … that he steels himself on pain of risking a loss of identity.[17]

Korsgaard herself, however, observes that "different laws hold for wantons, egoists, lovers, and Citizens of the Kingdom of Ends. In order to establish that there are ways in which we *must* think of our identities, and so that there are *moral* obligations, we will need another step."[18]

Korsgaard makes this additional step by appealing to the *necessity* of having an identity. Various conceptions of the self are themselves contingent: "You may cease to think of yourself as a mother or a citizen or a Quaker; or, where the facts make that impossible, the conception may cease to have practical force: you may stop caring

---

[16] Korsgaard, *The Sources of Normativity*, 113.

[17] Cohen, 183.

[18] Korsgaard, *The Sources of Normativity*, 113.

whether you live up to the demands of a practical role."[19] Nevertheless, "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 yourself as having any reason to live and act at all."[20] She continues,

> But *this* reason for conforming to your particular practical identities is not a reason that *springs from* one of those particular practical identities. It is a reason that springs from your humanity itself, from your identity simply as *a human being*, a reflective animal who needs reasons to act and to live. And so it is a reason you have only if you treat your humanity as a practical, normative, form of identity, that is, if you value yourself as a human being.[21]

Practical identities, then, depend on our valuing humanity.[22] It is from this valuing of humanity that our moral obligations arise.

In her response to Cohen's critique, Korsgaard claims that Cohen has failed to recognize the claims of humanity – that is, the claims of morality – on the Mafioso: "If Cohen's Mafioso attempted to answer the question why it matters that he should be strong and in his sense honour-bound even when he is tempted not to, he would find that its mattering depends on the value of his humanity, and … he would find that that commits him to the value of humanity in general, and so to giving up his role as a Mafioso."[23] But I think this response fails to appreciate the force of Cohen's critique.

---

[19] Ibid., 120.

[20] Ibid., 120-121.

[21] Ibid., 121.

[22] The careful reader will notice that I have shifted from talking of valuing one's *own* humanity to valuing humanity *as such*. Korsgaard addresses this issue in Lecture 4 of *The Sources of Normativity*. I will not go into her solution here, as I want to focus on other features of her position, but it will come up in chapter three.

[23] Korsgaard, *The Sources of Normativity*, 257. It is worth noting, too, that Korsgaard maintains that the Mafioso *does* have an obligation to fulfill his role as a Mafioso insofar as he continues to

Cohen observes that "the mafioso can honour human beings the springs of whose actions are congruent with his *own* practical identity."[24] I take it that Cohen is trying to suggest that the practical identity of the Mafioso determines what it means to value humanity. Why couldn't the Mafioso claim that in using the means necessary to realize his ideals he *is* valuing humanity precisely because he is encouraging himself and others to achieve this ideal? I am not sure this is correct, though I am sympathetic to the idea, but I do think Cohen points us towards a question Korsgaard needs to answer: what does it mean to value humanity?

Before considering what Korsgaard has to say about what it means to value humanity, it is worth addressing the concern that I might be changing the subject on Korsgaard. After all, she is attempting to answer the normative question which is posed *after* the content of morality has been established. By asking her to fill out what it means to value humanity, am I asking her to provide the content of morality rather than challenging her response to the normative question? I think not. First of all, if the fact that we must 'value humanity' is supposed to account for why we must be moral, there must be some connection between valuing humanity and moral demands. Secondly, Korsgaard's constructivist project is to explain why we must be moral by showing how moral demands follow from proper reflection. If we can justify *not* being moral while adhering to her account of proper reasoning, her response to the normative question will have failed. Asking for an account of what it means to value humanity may also involve

---

reflectively endorse this way of life. It is simply that these obligations conflict with his moral obligations, and if he were to live consistently he would be constrained to abandon his life as a Mafioso.

[24] Cohen, 187.

asking for an account of the content of morality, but Korsgaard must provide this account to achieve her goal of answering the moral skeptic.

What, then, does Korsgaard think is involved in 'valuing humanity'? Her appeal to 'valuing humanity' is inspired by Kant's second formulation of the categorical imperative: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means."[25] I propose, then, to look to Korsgaard's discussion of this formulation of the categorical imperative, the Formula of Humanity, to uncover her thoughts on what it means to value humanity.

Citing a passage in *The Metaphysics of Morals* where Kant claims that "[t]he capacity to set oneself an end – any end whatsoever – is what characterizes humanity (as distinguished from animality),"[26] along with several other similar passages, she concludes that "[i]t is this capacity that the Formula of Humanity commands us never to treat as a mere means, but always as an end in itself."[27] But what does it mean to treat the capacity to set ends for oneself as an end in itself? Normally, when we think of treating something as an end we think of trying to bring it about, but the capacity to set ends for oneself is not something we try to bring about in ourselves or others. Rather, Korsgaard maintains, to treat this capacity as an end in itself is to realize that "it must *never* be acted against."[28] In other words, we must never undermine the capacity of anyone to choose for herself. Thus, Korsgaard suggests elsewhere that, in our treatment of others, "an action is

---

[25] Kant, *G.*, 4:429.

[26] Kant, *MS*, 6:392.

[27] Christine M. Korsgaard, "Kant's Formula of Humanity" in *Creating the Kingdom of Ends* (New York: Cambridge UP, 1996): 106-132. See p. 114.

[28] Ibid., 125, emphasis hers.

contrary to perfect duty if it is not possible for the other to assent to it or to hold its end."[29]

She emphasizes that it must be the case not that the other *would* not assent but that the other *cannot* assent to the action. Even if the other would not assent, so long as it is possible for him to assent or dissent, his ability to set his own ends is not undercut. This interpretation of what it is to respect the humanity of others effectively rules out coercive and deceptive actions.[30] Coercive actions, of course, deny the other person the opportunity to assent or dissent. And since deceptive actions by their nature depend on the other person's ignorance of the deception, they also deprive the other person of the opportunity to assent or dissent. Perhaps, then, Korsgaard can effectively prohibit the actions of the Idealized Mafioso since, whatever his motives, he certainly employs coercive and deceptive tactics. However, this interpretation of the Formula of Humanity is problematic.

Japa Pallikkathayil has raised several puzzles for Korsgaard's take on the Formula of Humanity.[31] In particular, she has pointed out that this interpretation of the second formulation of the categorical imperative rules out too much: coercion and deception are

[29] Christine M. Korsgaard, "The Right to Lie: Kant on Dealing with Evil" in *Creating the Kingdom of Ends* (New York: Cambridge UP, 1996): 133-158. See p. 138.

[30] In another article, Korsgaard claims that this interpretation rules out force, coercion, and deception. See Christine M. Korsgaard, "The Reasons We Can Share: An Attack on the Distinction between Agent-Relative and Agent-Neutral Values" in *Creating the Kingdom of Ends* (New York: Cambridge UP, 1996): 275-310. See p. 295. The distinction she makes between force and coercion is not necessary for my purposes, so I will regard the use of force as a kind of coercion.

[31] Pallikkathayill identifies both 'interpretative' and 'substantive' problems for Korsgaard's interpretation of the Formula of Humanity. The interpretative problem contends that Korsgaard's view does not accurately represent Kant's own. The substantive problem raises objections to Korsgaard's view in its own right. I will focus on the latter. Japa Pallikkathayill, "Deriving Morality from Politics: Rethinking the Formula of Humanity," *Ethics* 121 (October 2010): 116-147.

not always wrong.[32] Though Kant notoriously denies that it is permissible to lie to the murderer at the door, most people believe he is mistaken. At the very least, it is surely permissible forcibly to restrain the murderer. And Kant himself grants that we can coerce others to enter into a civil society with us.[33] But the problem is worse yet for Korsgaard. Pallikkathayil observes that the motivation for insisting that consent must be possible actually commits Korsgaard to a stronger criterion: consent must be meaningful.

Recall that Korsgaard's claim that we should not treat others in ways to which they cannot possibly assent stems from her commitment to the idea that respecting humanity means not hindering the capacity to set ends for oneself. She expands on this idea: "The idea of deciding for yourself whether you will contribute to a given end can be represented as a decision whether to initiate that causal chain which constitutes your contribution. Any action which prevents or diverts you from making this initiating decision is one that treats you … as a mere means, a thing, a tool."[34] If I do something to you to which you *can* assent, but intend to proceed with my course of action regardless of whether you assent to it, I rob you of the ability to make this 'initiating decision.' Even though your assent is possible, it means nothing to me and contributes nothing to my decision to treat you in a certain way.

In some cases, Pallikkathayil is happy to admit that my failure to take your assent or dissent into account reveals something problematic in my action. If I ask you if I may have the last piece of cake, but then proceed to take it even after you have told me not to,

[32] As we will see in chapter three, Korsgaard admits as much herself. I will expand on the problems with this admission in that chapter.

[33] Cf. Kant, *MS*, 6:256 and 6:264.

[34] Korsgaard, "The Right to Lie," 140-141.

I do show a lack of respect for you. However, it cannot be the case that a lack of meaningful consent is sufficient to make an action wrong. Pallikkathayil suggests a case in which a person looks out the window to determine whether she needs to wear a coat by observing the dress of passersby. It is not impossible for the passersby to assent or dissent to being observed, but the fact of the matter is that the onlooker does not *care* whether they assent or dissent. And there doesn't seem to be anything morally problematic about this attitude on the part of the onlooker.[35]

The significance of these challenges is that Korsgaard's interpretation of the Formula of Humanity does not provide a satisfying account of what it is to value humanity. Some other account must be provided. In particular, insofar as it is plausible that what it means to value humanity depends on one's practical identity, we need an account which can rule out the Idealized Mafioso. Unfortunately, it is not clear to me that Korsgaard's constructivism will allow her to provide an alternative account of what it is to value humanity. After all, her constructivism commits her to the view that our capacity to set ends "is the condition of the goodness of all our other ends."[36] And this, she says, "in a sense gives the reason"[37] why humanity is never to be acted against. I cannot conclude that it would be impossible for Korsgaard to provide an alternative account of what it means to value humanity. However, if our capacity to choose is the source of *all* value, it seems that the only criterion for determining whether an action fails to treat humanity appropriately is that it undercuts this capacity. If that is the only

[35] Pallikkathayil, 127.

[36] Korsgaard, "Kant's Formula of Humanity," 125.

[37] Ibid.

criterion, I cannot imagine what other account Korsgaard could give than the one discussed above. And, as we have seen, we have reason to reject that account.

Korsgaard's more recent work suggests a slightly different response to the moral skeptic which, I will argue, also fails, but which also reveals that Korsgaard's theory is *unable* to address the Idealized Mafioso. In fact, it cannot even acknowledge the possibility of his existence, let alone persuade him to live a morally upstanding life. Korsgaard still appeals to our identity as human beings, but ties normativity more explicitly to the demands of agency and the inescapability of action. Human beings are animals that must *act*, and it is this aspect of our identity that grounds moral obligation.

Korsgaard's Later Work: Self-Constitution

Action plays a special role in Korsgaard's most recent book, *Self-Constitution*. We not only *must* act, but in acting we constitute ourselves. Korsgaard explains, "When you deliberately decide what sorts of effects you will bring about in the world, you are also deliberately deciding what sort of a cause you will be. And this means you are deciding who you are."[38] But insofar as we are agents who choose how to act, we are required to act *as* agents, to constitute ourselves *as* agents. And Korsgaard maintains that "it is essential to the concept of agency that an agent be unified. That is to say: to regard some movement of my mind or my body as *my action*, I must see it as an expression of my self as a whole, rather than as a product of some force that is at work *on* me or *in* me. Movements that result from forces working *on* me or *in* me constitute things that happen to me."[39] In other words, when you act you establish yourself as the cause of your action.

[38] Korsgaard, *Self-Constitution*, 19.

[39] Ibid., 18.

But in order to establish *yourself* as the cause of your action, your action must proceed from and establish you as a unified whole so that the action is not merely the effect of one or more forces working on you. Thus, whatever the various ends of your different actions, it is the function of action as such to unify the agent.

Korsgaard believes this unity is achieved through adherence to the categorical imperative, which requires that any maxim behind an action be universalizable. That is, it must be possible to will the same maxim in all relevantly similar cases. The alternatives to willing a maxim universally are what Korsgaard calls 'particularistic willing' and willing a maxim 'in general.' Particularistic willing consists in "willing a maxim for exactly this occasion without taking it to have any other implications of any kind for any other occasion."[40] In other words, you take the means necessary to realize whatever incentive has a hold of you at the moment, and as you act, you withhold the will to act in a relevantly similar way in relevantly similar future circumstances. The next moment, when that incentive no longer has hold of you, you take on a course of action meant to realize the incentive of that moment, regardless of its relation to your previous or subsequent behavior and desires.

The problem with this sort of willing is that it does away with the distinction between the self and its incentives. Remember that in order to establish yourself as the cause of your action you must establish yourself as a unified whole. If your behavior is the result of whatever impulse or force grabs hold of you in the moment, it is not an action that *you* have brought about. Korsgaard believes that "when you deliberate, … it is as if there is something over and above all your incentives, something which is *you*,

[40] Ibid., 75.

and which chooses which incentive to act on."[41] Particularistic willing, however, requires you to "wholly identify with the incentive of your action. That incentive would be, for the moment, your law, the law that defines your agency or your will."[42] Thus, *you* never act. *You* never choose. *You* are reduced to a mere series of forces as each successive incentive or impulse demands your allegiance.

It might be tempting to think that the argument against particularistic willing does not commit one to willing universally, as the categorical imperative demands, but only to willing 'general rules.' Korsgaard believes this conclusion is incorrect, drawing on a distinction between a principle that holds 'in general' and one that is 'provisionally universal.' The former "applies to a wide range of similar cases;" the latter "applies to every case of a certain sort, unless there is some good reason why not."[43] Korsgaard adds, "Treating principles as general, and treating them as provisionally universal, seem superficially similar, because in both cases we admit that there might be exceptions."[44] But in fact there is an important difference, revealed by the fact that when we encounter an exception to the general principle nothing happens, but when we encounter an exception to a provisionally universal principle "we must now go back and revise [the principle]."[45] In other words, a general principle doesn't have the unifying effect Korsgaard wants. A provisionally universal principle, by contrast, aims at 'absolute universality.' It seeks to unify the agent across different circumstances and experiences.

---

[41] Ibid.

[42] Ibid.

[43] Ibid., 73.

[44] Ibid.

[45] Ibid., 74.

One might 'act' with the idea that one will so act 'generally' in the future, or one might act with the idea that one will always so act. Only the latter qualifies as taking oneself seriously as an agent and thus as genuinely *acting*.

Borrowing from Aristotelian ideas, Korsgaard is able to connect the categorical imperative with morality at this point. The function of action, remember, is to unify the agent who is acting. Bad action, Korsgaard concludes, fails to contribute to the constitution and maintenance of the agent. A good action, by contrast, is good *qua* action if it contributes to the unification of the agent. But to judge an action as good or bad *qua* action is to make a moral judgment. In a similar way, Korsgaard connects morality with autonomy. To be autonomous is just to determine oneself to action rather than to be determined by outside impulses. If a good action is good insofar as it contributes to the unification of the agent, an action is good insofar as it contributes to the realization of autonomy.

Thus, Korsgaard has argued that autonomy is a *constitutive* standard of *action*. A constitutive standard of a thing is such that failure to meet that standard is a failure to be the thing in question. For example, a house must be constituted in certain ways in order to count as a house. An action must contribute to the constitution and maintenance of autonomy to count as an action. The nice thing about standards of this sort is that they "meet skeptical challenges to their authority with ease."[46] Korsgaard explains, "Because it does not make sense to ask why a house should serve as a shelter, it also does not make sense to ask why the corners should be sealed and the roof should be waterproof and tight. For if you fall too far short of the constitutive standard, what you produce will

[46] Ibid., 29.

simply not be a house."[47] Since human beings must act, and since action must contribute to the autonomy of the agent to count as action, Korsgaard believes she has the tools necessary to respond to skeptics who doubt that they need to be moral.

But why? Clearly some builders produce shoddy houses. What reason do we have *not* to produce shoddy actions? The answer lies in Korsgaard's understanding of what happens in the case of the builder who produces shoddy houses. "The shoddy builder," says Korsgaard, "doesn't follow a different set of standards or norms. He may be doing one of two things. He may be guided by the norms, but carelessly, inattentively, choosing second-rate materials in a random way, sealing the corners imperfectly, adding insufficient insulation, and so on."[48] Alternatively, however, she concedes that he might "be doing this sort of thing quite consciously, say in order to save money."[49] Such a person is not, according to Korsgaard, trying to build a good house. Rather, she thinks that "we should follow Socrates's lead, and say that he is not trying to build a house at all, but rather a sort of plausible imitation of a house, one he can pass off as the real thing. What guides him is not the aim of producing a house, but the aim of producing something that will fetch the price of a house, sufficiently like a real house that he can't be sued for it afterwards."[50]

I want to bracket the question of whether Korsgaard is right in her conviction that immoral action is explained by its failure to meet the constitutive standards of action – for now – and assess whether this appeal to so-called constitutive standards of action

[47] Ibid.

[48] Ibid., 31.

[49] Ibid.

[50] Ibid., 31-32.

adequately responds to the challenges posed by moral skeptics. David Enoch has argued that it does not. In his article, "Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Action," he imagines a moral skeptic who responds to Korsgaard thus:

> If your reasoning works, this just shows that I don't care about agency and action. I am perfectly happy being a shmagent – a nonagent who is very similar to agents but who lacks the aim (constitutive of agency but not of shmagency) of self-constitution. I am perfectly happy performing shmactions – nonaction events that are very similar to actions but that lack the aim (constitutive of actions but not of shmactions) of self-constitution.[51]

In response to Korsgaard's contention that action is unavoidable, he writes, "'Perhaps,' Korsgaard's skeptic might say, 'I cannot opt out of the game of agency, but I can certainly play it half-heartedly, indeed under protest, without accepting the aims purportedly constitutive of it as mine."[52]

This argument, I believe, gets something right. Consider again the example of the house. Enoch wants to say that someone can be convinced that there are certain constitutive standards the builder is not meeting in building a house, but that he can respond by saying, "Very well then, I guess I am not engaging in the project of building a house but rather in the project of building a shmouse, of which these standards aren't constitutive. So what is it to me how you classify my project?"[53] To the suggestion that we do not *have* to build houses but that we *have* to act, he asks why we must act *well*.[54]

---

[51] David Enoch, "Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Action," *Philosophical Review* (2006): 169-198.

[52] Ibid., 188.

[53] Ibid., 179.

[54] I wonder whether Enoch could go farther than he does and simply deny that Korsgaard is right to insist that we must act. If Korsgaard's argument that all action is morally good works, her claim that we

Notice, however, that Korsgaard's further identification of good action as action that contributes to the constitution and maintenance of the self entails that the question of why we must act well boils down to asking why we should aim for the greatest degree of unity within the self. Why not act to retain a minimal amount of unity if doing so will, say, promise more pleasure of a certain kind? Korsgaard's response to this question will bring us back to the question of whether the demands of autonomy sufficiently account for morality.

Korsgaard maintains, plausibly enough, that "deliberative action by its very nature imposes unity on the soul. When you deliberate about what to do and then do it, what you are doing is organizing your appetite, reason, and spirit, into a unified system that yields an action that can be attributed to you as a person. Whatever else you are doing when you choose a deliberative action, you are also unifying yourself into a person."[55] Since just action contributes to this unification of the agent, while unjust action tends towards the disintegration of the person, Korsgaard believes the just life is the only one you *can* choose. After all, she argues, "you can't, in the moment of deliberative action, choose to be something less than a single unified agent. And that means you can't exactly choose to act on any principle other than the principle of justice."[56] Her argument might be summarized thus: Human beings must act, and the function of deliberative action is to unify the agent. However, only just actions succeed in bringing about that unification, so humans must rationally choose to act justly.

---

must act seems less intuitively plausible. After all, we all know people who do bad things. I will not pursue that idea here, however.

[55] Korsgaard, *Self-Constitution*, 179.

[56] Ibid., 180.

The Problem Again

But this response to the moral skeptic reveals that Korsgaard's failure to address the Idealized Mafioso is no mere oversight. The fact of the matter is that Korsgaard cannot make sense of the Idealized Mafioso precisely because the very *idea* of him is incoherent to her. There is no such thing as deliberately chosen immoral action for Korsgaard. Consider, after all, her claim that "you can't, in the moment of deliberative action, choose to be something less than a single unified agent. And that means you can't exactly choose to act on any principle other than the principle of justice."[57] In other words, when you choose to act on the basis of some reason you cannot at the same time choose to act in a way that does not meet the standards of morality as demanded by human agency. It has been recognized before that constructivist ethics reduces all immorality to irrationality,[58] but for the first time Korsgaard has explicitly reduced all immorality to a lack of agency.

Korsgaard recognizes this herself, however, affirming that "the unjust *person* cannot act at all,"[59] leading her to raise the question: "But then how is it possible for action to be defective, and still be action?"[60] I will discuss her response in more detail below, but she basically offers a theory in which "responsibility in general is going to

[57] Korsgaard, *Self-Constitution*, 180. I will comment on what she means by saying that a person cannot 'exactly' choose to act on any other principle than the principle of justice later.

[58] For example, Patrick Kain writes, "To violate a duty, … an agent would have to choose to obey the very same moral law he chooses to violate, since the former volition is, *ex hypothesi*, a necessary condition for the existence of the duty and the latter is required for the performance of the immoral action. On this interpretation, a will would have to be actually self-contradictory in order to violate duty." Patrick Kain, "Self-Legislation in Kant's Moral Philosophy," *Archiv für Geschichte der Philosophie* 86, no. 3 (2004): 257-306.

[59] Korsgaard, *Self-Constitution*, 152.

[60] Ibid., 161.

look a lot more like responsibility for omission."[61] She explains, "What we are going to blame you for is not that other force that was working in you or on you, but for the fact that you let it do that."[62] In other words, the only sense in which you are responsible for immoral action is that you allowed some impulse to overtake you – not that you deliberately *chose* to act immorally. The Mafioso is unaccounted for.

It might be tempting to think that this is as it should be. Perhaps deliberate, immoral action is simply unintelligible, nonsensical. After all, if we take Korsgaard's project of answering the moral skeptic seriously, isn't this precisely the outcome we should expect? If we have most reason to act morally, it follows that it is irrational to act immorally. Perhaps the great strength of Korsgaard's view is that it reveals the skeptic's position for the nonsense that it is.

I think this argument turns on a familiar ambiguity in what we mean by 'reason.' Consider Bernard Williams' famous example of the person who desires to take a drink from a bottle, believing it to contain gin. Unbeknownst to him, however, the bottle actually contains petrol. In some sense, given his beliefs and desires, he has reason to take a drink from the bottle. Let us call his reasons to take a drink *subjective* reasons, though in calling them 'subjective' I do not want to suggest that these reasons are somehow not 'real.' After all, the agent is not irrationally compelled to take a drink but reasonably determining, in light of his beliefs, that taking a drink is a good idea. But in another sense, most of us would agree that he does not have reason to take a drink from the bottle. In fact, we would want to maintain that he has reason *not* to take a drink from

[61] Ibid., 175.

[62] Ibid.

the bottle. Let us call these reasons *objective* reasons.[63] Similarly, I want to affirm that just as this person has reason not to take a drink from the bottle, we have objective reason – even overriding reason – to act morally. However, just as this same person also has a reason in another sense *to* take a drink from the bottle, we can have subjective reason to act immorally.

But if our account of deliberate immoral action requires us to maintain that the person who chooses to act immorally does so as the result of a mistake, just as the person who chooses to take a drink out of the bottle does so as the result of the mistaken belief that the bottle contains gin rather than petrol, we might worry that it will leave us unable to attribute responsibility correctly. After all, if someone deliberately chooses to do the wrong thing as the result of a mistake, don't we excuse that person rather than hold her responsible?

But this argument also plays off of an ambiguity. The mistake in judgment that the wrongdoer makes is a mistake about what she has most reason to do – not about what morality demands of her. Recall that Korsgaard first poses the normative question as a question that arises *after* the content of morality has been established. That is, we can have some understanding of the demands of morality without understanding that reason requires us to act accordingly.[64] A person may judge that something is morally wrong

[63] I take this example from Bernard Williams, "Internal and External Reasons" in *Moral Luck* (New York: Cambridge UP, 1981), 101-113. See pp. 102-103. What I call 'objective' and 'subjective' reasons might map onto what Williams calls 'external' and 'internal,' respectively. I want to avoid addressing debates about the existence of external reasons for now. Williams seems to intend the gin/petrol case to deal strictly with internal reasons, and what I call an 'objective reason' might at this point be interpreted as what Williams calls 'an optimistic internal reason claim' rather than an external reason. Williams, 111.

[64] Elsewhere Korsgaard makes a distinction between substantive and procedural justice that may provide a helpful illustration of what I have in mind. A court that renders a decision that lets a guilty person go free may have rendered the just verdict if it followed the correct process in arriving at its decision

and yet decide that he has more reason to do it than not to do it. For example, a naïve agent who has not thought through the basis of moral judgments might simultaneously believe "It is wrong to steal" and "I have most reason to steal."[65] Even more reflective agents might think something like "It is wrong to steal because I would resent it if someone stole from me, and there is no good reason to demand special treatment for myself" while also thinking "I have most reason to steal." Thus, in the terminology above, a person may rationally believe he has most objective reason to do something immoral – and so have most subjective reason to perform that immoral act – and be mistaken. To make sense of deliberate immoral action, we must maintain a conceptual difference between the content of morality and our reasons for acting accordingly, even if reason ultimately requires us to act morally.

To some extent, Korsgaard should be comfortable with this acknowledgement: just as it makes sense to ask whether water is H_2O, it may make sense to ask whether we have reason to act morally even if she is correct in ultimately identifying morally permissible action with what we have reason to do. Besides, it is only by maintaining a conceptual difference between morality and what we have most reason to do that Korsgaard can motivate her project. How else could she think *both* that the moral thing to do is the thing we have most reason to do *and* that it makes sense to ask why we

---

– that is, it might be *procedurally* just – even though most of us would regard letting a guilty person go free *substantively* unjust. See Christine M. Korsgaard, *Self-Constitution*, 148-149. However, Korsgaard thinks the only way to arrive at substance is through correct procedure. That is, moral content depends on the correct rational procedures of agents. In *The Sources of Normativity*, she therefore tells us that she is a procedural realist rather than a substantive realist: there is moral truth because there are correct procedures for arriving at it, not because there are substantive moral facts out there in the world to which the moral truth corresponds. See Korsgaard, *The Sources of Normativity*, 35.

[65] I have in mind Aristotle's example of the person who knows the "that" of ethics before he knows the "why." See Aristotle, *Nicomachean Ethics*, trans. Terence Irwin (Indianapolis, IN: Hackett, 1999), 1095b.

should do the moral thing? How else could she make the normative question the focal point of her project? Moral skepticism could not pose a genuine threat without this conceptual distinction in place. But in addition to making sense of the normative question, maintaining this conceptual difference also makes room for the possibility that deliberate immoral action might be intelligible.

And this presents a problem for Korsgaard's answer to the normative question. It might make sense to ask what reason we have to be moral, just as it makes sense to ask whether water is H_2O, but to respond by saying that acting morally is simply what it *is* to act – that it is the only way we *can* rationally act – is to rule out the possibility of deliberate immoral action. While there is something appealing about the Socratic idea that deliberate immoral action is nonsensical, we frequently interpret people as deliberately choosing to act wrongly – not as simply allowing themselves to be governed by some impulse or another. And in light of the failure of the foregoing argument that deliberate immoral action is unintelligible, our tendency to interpret ourselves and others as sometimes choosing to act wrongly cannot be simply dismissed. As Candace Vogler points out in the quotation that begins this chapter, our theories of action and morality begin with our observations of people in everyday life. We need an account of ethics that enables us to make sense of deliberate immoral action precisely because deliberate immoral action *is* a conceptual possibility for us. Immoral actions *as actions* are not unintelligible to us, and any theory that renders them such is thereby defective.

Indeed, we are now in position to start sketching an account of what must be the case if we are to make sense of deliberate immoral action. If Korsgaard's project of responding to the moral skeptic is to succeed, the agent needs to have conclusive

objective reason to act in accordance with what morality demands. For a person to have subjective reason to act immorally, then, he must be mistaken about his objective reasons. Korsgaard can, to some extent, accept this ambiguity in how we talk about reasons for action. She can, for example, grant that there are subjective reasons for other sorts of mistaken behavior, like taking a drink from the bottle of petrol. But those sorts of actions won't be immoral – just silly or regrettable. She can even describe the person who 'acts' immorally as mistakenly taking himself to have reasons for his behavior. But that is the problem: Korsgaard must say that the person who 'acts' immorally mistakenly takes himself *to have reasons*.

According to Korsgaard, what makes immoral behavior immoral is that the agent acts on an incentive that cannot possibly be regarded as a reason for action. It is not simply that the agent acts on an incentive that cannot *be* a reason for action; rather, the agent cannot even reasonably *regard* the incentive as a reason. There can be no subjective reason for immoral action. If the agent follows the procedures of deliberation correctly, she will discover what morality requires of her *by* realizing what she has reason to do. Though I have not yet filled in the details of Korsgaard's position, recall that failure to follow the procedures of deliberation correctly is to be overcome by some desire rather than to choose to act on a reason. For an immoral action to be the object of choice, we need it to be the case that the agent genuinely *has reasons* for his action but that he is mistaken about what he has *most* or *conclusive* reason to do.[66] Korsgaard cannot, however, admit that there could ever be subjective reason to act in a way that one

[66] Vogler seems to have a view like this in mind when she contends that contemporary ethicists could benefit from studying Thomas Aquinas's work on vice. In any case, it seems clear that Aquinas endorses the claim that those who choose to act wrongly have reasons for their actions, but they are mistaken about what they have conclusive objective reason to do. See Aquinas, *DM*, Q.III.12.

knows to be immoral. In order for deliberate immoral action to be possible, however, it must be possible for the agent to choose to act in that way, i.e., to have a subjective reason to act immorally.

When it comes to moral reasoning, Korsgaard's constructivism commits her to collapsing the distinction between objective and subjective reasons. If our reasons are constructed by following correct deliberative procedures, then objective reasons are always subjective and subjective reasons are always objective. It is certainly possible for a person to be driven to act on some incentive she mistakenly takes to be a reason – but in that case her reasoning is overwhelmed by that incentive or impulse. It is not the case that she functions as a rational agent who chooses to act on a subjective reason. And this brings me to a further point: making sense of deliberate immoral action pushes us toward adopting some form of moral realism. If making sense of immoral action requires that there be objective reasons about which we can be mistaken, it seems to follow that those reasons must exist independently of the human mind. The need for an account of immoral action does not simply lead us to abandon constructivism; it gives us reason to embrace realism.

Another Problem

Furthermore, not only can we make sense of the idea of deliberate immoral action, but to render deliberate immoral action unintelligible by claiming that it does not count as genuine action is to compromise an important feature of ethical living, one that is especially important to Kantians. Namely, if all immoral behavior is regarded as a *lack* of agency, a basic respect for others that features prominently in our interactions with them – including those who do wrong – is undermined.

Recall that Korsgaard has argued persuasively that Kant's call to respect humanity, in ourselves and others, is a call to respect the capacity to *choose*, to set ends for oneself. It is interesting, then, that later in *The Metaphysics of Morals* Kant insists, "I cannot deny all respect to even a vicious man as a human being; I cannot withdraw at least the respect that belongs to him in his quality as a human being, even though his deeds make him unworthy of it."[67] On the face of it, this comment of Kant's does not conflict with Korsgaard's view. After all, she maintains that that we can still hold people responsible for their wrongdoing precisely because they *allowed* themselves to be governed by some force working on them. This claim indicates that we think of wrongdoers as having a capacity *not* to be so governed – that is, a capacity to set ends for themselves – which they simply do not exercise. However, I think two points suggest that Kant means to say something that Korsgaard cannot account for.

First, in the "Remark" that immediately follows Kant's insistence that we still accord a basic respect to wrongdoers, he compares the respect we give them to the respect we show a person who makes an error in logical reasoning:

> On this is based a duty to respect a human being even in the logical use of his reason, a duty not to censure his errors by calling them absurdities, poor judgment, and so forth, but rather to suppose that his judgment must yet contain some truth and to seek this out, uncovering, at the same time, the deceptive illusion (the subjective ground that determined his judgment which, by an oversight, he took for objective), and so, by explaining to him the possibility of his having erred, to preserve his respect for his own understanding… The same thing applies to the censure of vice, which must never break out into complete contempt and denial of any moral worth to a vicious human being; for on this supposition he could never be improved, and this is not consistent with the idea of a *human being*, who as such (as a moral being) can never lose entirely his predisposition to the good.[68]

[67] Kant, *MS*, 6:463.

[68] Ibid., 6:463-464, italics in original.

This passage suggests that respecting the humanity of others includes not only assuming that they have the capacity to set ends for themselves but that they *are* in fact exercising this capacity, just as we assume that those who make errors in logic are nonetheless exercising their reason.

Indeed, charity requires us to assume that they exercise this capacity lest we rule out the possibility of their improvement. If we think that in doing wrong they fail to exercise their capacity to choose ends for themselves, even momentarily, we give up hope for their improvement – we write them off. After all, even if it is just a momentary failure to exercise practical reason, no amount of reasoning with them can possibly prevent them from failing to reason in the future. They are like animals or children who may at any time be overcome by some passing impulse. But though you cannot teach someone who does not reason to reason, you can teach someone who reasons poorly to reason better.

Thus, just as we assume that someone who makes an error in judgment is exercising his reason even though he errs, so we should assume that the vicious person is exercising her reason even though she errs. But again, this seems to be precisely what Korsgaard *cannot* say. After all, the person who behaves immorally on her account is not acting for a reason but is overwhelmed by sensible impulses. It is not that the person who behaves immorally acts on a reason but is simply mistaken about what she actually has most reason to do. Rather, she acts for no reason at all, impelled by forces she cannot identify with herself as a unified agent. No amount of reasoning with such a person will prevent her from being impelled by forces in the future.

The second point concerns the nature of responsibility or imputation. While Korsgaard preserves a space for responsibility, her account seems strikingly different from the account proffered by Kant who defines imputation as "the *judgment* by which someone is regarded as the author (*causa libera*) of an action."[69] Notice that imputation, on this view, is a way of showing respect for a person because imputing an action to a person is to recognize that he chooses to perform that action. And Kant clearly believes we can impute immoral action to people. But on Korsgaard's account, someone who does wrong is *never* the author of an action but is merely overwhelmed by some external force. Thus, the action cannot be imputed to the wrongdoer in Kant's sense: the wrongdoer is merely caught in the throes of his passions.

But in that case anyone who commits wrong seems comparable to a child caught in a fit of ill-temper: capable, perhaps, of exercising reason, but ultimately allowing himself to be overwhelmed by his impulses. Our reactions to wrongdoers should then be comparable to our reactions to children throwing temper tantrums. Whether or not Kant's ethics commits him to the view that respect for the humanity of others requires that we assume that wrongdoers in fact *choose* to perform immoral actions, I think many of us would agree that there is something wrong with regarding all wrongdoing as akin to the temper tantrums of children. Not only is it simply not true that we regard all wrongdoers this way (though it may be how we regard some), there seems to be something objectionably paternalistic about such a view. One could understandably take offense at being treated as if one were somehow not 'in control' when one acted wrongly.

[69] Ibid., 6:227, italics in original.

Solutions?

Korsgaard of course acknowledges that people seem to act badly, so she does attempt to offer some account of bad action. Recall that she compares acting badly to badly building a house: to act badly is just to fail to conform completely to the constitutive standards of acting just as badly building a house is to fail to conform completely to the constitutive standards of house-building. Given that she identifies the constitutive standard of action with the ability of the action to contribute to the constitution of the self into a unified whole, she identifies bad action with action that contributes in some way to the self's disintegration. In other words, just as badly building a house leads to the construction of a structure that is poorly made, bad action leads to a 'bad constitution' of the self. Thus, to explain the relationship between bad action and the agent, she attempts to outline what a 'bad constitution' might look like.

To begin, let me say a little more about what Korsgaard thinks the successfully unified self looks like. Korsgaard expands her idea of autonomy, or self-rule, describing it as the result of each part of the soul playing its proper role, much as Plato describes in the *Republic*. The self is not identified with reason, passion, or appetite, but with the proper constitution of these parts into a unified whole. Reason rules because that is its proper function, not as one force that merely overpowers another. The appetites suggest certain ends to reason, and reason considers which of these ends it can bring into being through the execution of a universalizable law. Bad actions, then, are actions that do not contribute to or proceed from the self understood in this way. In fact, Korsgaard gives an overview of four 'bad constitutions' and the kinds of actions associated with them, based once again on Plato's four bad constitutions. She believes these bad constitutions may

give a human being some semblance of a unified life for a while, but ultimately they tend towards the disintegration of the self. I take it that this is why she says that a person cannot 'exactly' choose to act wrongly: because there is a sense in which behavior that does not perfectly unify the agent may yet contribute to a constitution that gives the agent some minimal degree of unification.[70] This behavior thus looks superficially similar to genuine, chosen action. My question is whether all immoral behavior can be explained as contributing to the formation of a bad constitution in Korsgaard's sense.

I am willing to admit that autonomy understood as Korsgaard describes it may manage to rule out some immoral – or at least undesirable – behavior. Korsgaard's description of the democratic person is an excellent example of a kind of flawed life that is ruled out by her account of autonomy. This person, says Korsgaard, "is kind of wanton."[71] He indiscriminately pursues whatever desires happen to arise in him. As a result, "the coherence of the democratic person's life is completely dependent on the accidental coherence of his desires."[72] And that is precisely the problem with this way of life, "for it is … only an accident if [incoherence] does not happen. The democratic person has no resources for shaping his will to prevent this, and so he is at the mercy of accident."[73] Thus, the lack of autonomy does appear to explain what is going wrong in this person's life. But notice that the kinds of 'actions' that lead to the defectiveness of

[70] Alternatively, I think she might be referring to the fact that the person who does wrong 'allows' himself to be led by a desire or impulse. This 'allowing' might be construed as a choice, but it is not the kind of choice I want to make room for since it would still require us to interpret particular immoral 'actions' as the result of overwhelming desires rather than rational choice.

[71] Korsgaard, *Self-Constitution*, 168.

[72] Ibid., 169.

[73] Ibid.

the agent's constitution are the result of his submission to his impulses – not his reflective choice to act in a particular way. Is there a bad constitution that results from someone deliberately choosing to act wrongly?

Korsgaard comes closest to addressing this question in her discussion of the tyrannical constitution. On the one hand, she says, we think of evil as a privation. According to this view of evil, "bad people are people without standards, without integrity, without plans even, who can be led in any direction by the desire or the suggestion of the moment."[74] "But then," she asks, "there is that other vision, isn't there?"[75] She continues,

> According to this view, the bad or evil person is powerful, ruthless, unconstrained. The evil person is prepared to do *whatever is necessary* to get what he wants, and determined to let nothing stand in his way. … The tyrant of the ancient Greek imagination is the glamorous mafia kingpin of our own. So far from being *unable* to sustain relationships or projects, the evil person is more than anybody else able to stay on the track of them. For he is the one who is prepared to do *whatever is necessary*, whatever it takes. And this is where the doubt about morality comes in.[76]

Her response to this view of evil, quite simply, is that the tyrant, though unified, is not self-governed. He is a slave. According to Plato, the tyrannical soul is dominated by a single erotic desire, and Korsgaard elaborates this idea. Recall that Korsgaard defined an action as an act done for the sake of an end. Given this definition, Korsgaard concludes that the tyrant does not choose actions since "the tyrannical person doesn't choose *an act for the sake of an end*, the whole package as something worth doing. There's … one end or act that he's going to pursue or to do *no matter what*, and it rules him. And for him

---

[74] Ibid., 170.

[75] Ibid.

[76] Ibid., 170-171.

that end makes anything worth doing, anything at all, and that is a fact that is settled in advance of reflection."[77]

What we have here is an account that explains how immoral action can appear to follow from a unified will. It is certainly a plausible account. It sounds strikingly reminiscent of what Candace Vogler has to say about capital vice in her book *Reasonably Vicious*. Vogler writes that "capital vice is cultivated. One habituates oneself to it rather than being afflicted by it. It shows *strength*, not weakness, of will."[78] She even seems to regard capital vice as a kind of fixation on a single object: "Pride seeks the excellence of honor and renown. Avarice has as its object acquisition and control of wealth. Gluttony fixates on eating … Lust has sexual pleasure as its object."[79] But despite Korsgaard's apparent success in describing a particular form of vicious living, I do not think this account addresses the concern I am raising.

Korsgaard observes that she cannot logically prove that the tyrant is compelled by one overriding impulse, though in some cases – like that of the addict or the serial killer – it seems likely that some sort of compulsion rules over the agent. In itself, this is not a problem: she may not be able to prove that the tyrant is in the grip of some controlling desire, but that is her story of what must be going on in the case of the person who appears to have chosen an evil way of life. It follows from her theory of action. The trouble is that we can clearly imagine cases in which it would be difficult to maintain that the person is caught in the grip of some overwhelming impulse. Not only is there no

[77] Ibid., 172.

[78] Vogler, 38.

[79] Ibid., 39.

logical reason to suppose that someone who appears to choose a vicious way of life is in the grip of a desire, there is reason to think otherwise. Cohen's Idealized Mafioso springs readily to mind once again. It is not at all clear what single end would supposedly dominate the life of a Mafioso such that it could be said that his life is ruled by that end. Cohen describes him as living according to "a code of strength and honour."[80] This sounds like a life lived according to a certain principle rather than dominated by a single end. It seems much more reasonable to interpret the Mafioso as selecting certain ends and actions in keeping with a certain way of life that he has chosen.

But Korsgaard insists that the problem with the tyrant is that he "can never separate himself from *one* of his impulses, and so consolidates himself into a mere force of nature, an object, a thing."[81] And she may be right: this may be the problem with the tyrant. But in that case, the tyrannical soul does not provide an exhaustive model for how immoral behavior might appear to proceed from a deliberate, unified will. While this may be the problem with the type of soul she is describing, she has not yet accounted for the type of action I have argued she needs to explain – or at least explain away.

Given the frequency with which Cohen's Idealized Mafioso has appeared in this discussion, and given Cohen's claim that this Mafioso lives according to "a code of strength and honour," it might be worthwhile to consider the timocratic constitution – that is, the person in whom "spirit, the sense of honor, has usurped the role of reason."[82] According to Korsgaard, this person is taken with a sense of "moral glamour."[83] Indeed,

---

[80] Cohen, 183.

[81] Korsgaard, *Self-Constitution*, 173.

[82] Ibid., 165.

[83] Ibid.

the timocratic person cannot tell the difference between the appearance of goodness and goodness itself. Rather, she acts on the basis of what looks or feels right, of what is most in keeping with the grandeur she believes to accompany morality. I want to make three observations about the timocratic constitution.

First, the timocratic constitution also cannot account for deliberate immoral action. Korsgaard does not develop the idea of a 'spirited' part of the soul, but she does tell us that it "is a source for incentives."[84] In the timocratic constitution, the incentives provided by spirit overrule reason. Thus, the actions that proceed from the timocratic constitution are not reflectively chosen.

Perhaps it might be thought that nonetheless the timocratic constitution does accurately portray Cohen's Mafioso: he does not reflectively choose the way of life of the Mafioso but is overtaken by a kind of aesthetic appreciation for the glory of living up to certain ideals. I am not sure this is accurate. I think perhaps Cohen's concept of honor may have involved a commitment to a certain set of ideals rather than a commitment to the glory of living in accordance with those ideals. But no matter. Even if Cohen's Mafioso is captured by Korsgaard's discussion of the timocratic constitution, we can easily amend his example so that the Mafioso is committed directly to the values and ideals of life as a Mafioso. I will continue, then, to press Cohen's idealized Mafioso as a problem case for Korsgaard.

My second and third points are both related to Korsgaard's claim that the timocratic soul *is* unified for the most part. That 'for the most part' is, of course, important. Because the timocratic person loves the glory of moral living, her behavior

[84] Ibid.

typically conforms to what is required of her. Furthermore, insofar as she is committed across time to living morally – even if for the wrong reason – she achieves a kind of unity. The fact that she does not achieve total unity as an agent is revealed, Korsgaard thinks, only when the action that is called for is not what morality typically demands: "[The timocratic person] does fine, except in those moments when what is called for is concession, compromise, a bending of the rules, or even – as for instance in a case of civil disobedience – actions that are in some formal sense wrong."[85] I call our attention to this comment because, as I will argue in my third chapter, I think Korsgaard's suggestion that what is sometimes called for is "concession, compromise, a bending of the rules" creates problems for her view. But more on that later.

A third point, more interesting for my present purposes, follows more directly from her admission that the timocratic person is relatively unified. Though it may not be obvious from the discussion of the democratic and tyrannical constitutions, like Plato, Korsgaard thinks these constitutions form a hierarchy: some are more unified than others. In other words, agency comes in degrees. The more unified you are, the more of an agent you are. You may not embody an aristocratic constitution, but as a timocrat you're much more of an agent than the democrat. But if agency comes in degrees, this suggests another approach to responding to my challenge: perhaps we attribute immoral actions to others because they are more-or-less agents, even if they are not fully so.

But this move is not satisfactory. Remember, the reason that the person in question is only 'more-or-less' an agent – and not fully an agent – is that he sometimes allows himself to be led by other forces, whether desires or spirit. More specifically,

[85] Ibid.

whenever he 'acts' immorally he is being overwhelmed by some incentive not properly endorsed by reason. Thus, it remains the case that nobody deliberately chooses to do wrong. We simply yield to impulses and forces acting on us. And this brings me to my concluding point: no 'bad constitution' could *possibly* provide the account I demand of Korsgaard.

Korsgaard points out that she is not trying to provide a comprehensive catalog of bad constitutions, so one might think that perhaps there are other bad constitutions she has not identified that could account for the appearance of deliberate immoral action. But consider what it is for the self to have a bad constitution: it is for the parts of the soul *not* to serve their proper functions – that is, for reason to fail to rule, or for the desires to usurp the authority of reason.[86] Any account of a bad constitution, then, *must* involve the agent being overwhelmed by incentive rather than choosing to act in accordance with a reason. Even if Korsgaard can explain how bad action is consistent with a sort of unified 'will,' as in her discussion of the tyrannical and timocratic souls, it is in principle impossible for her to offer an account of someone *choosing* to act wrongly.

Korsgaard's discussion of bad constitutions merely confirms that she does not believe immoral actions can be chosen but are merely the result of people being overwhelmed by their impulses. Her attempt to respond to the skeptic by maintaining that "you can't, in the moment of deliberative action, choose to be something less than a single unified agent" [87] undermines any attempt to account for the possibility of

[86] Since Korsgaard treats the spirited part of the soul very much like the desires – that is, as offering incentives for action that need to be subjected to rational reflection and endorsement – I do not feel the need to address its role in the constitution of the soul.

[87] Korsgaard, *Self-Constitution*, 180.

deliberately chosen immoral actions. She still owes us an explanation, and it is not clear that she can give us one.

Korsgaard on Responsibility and Respect

Korsgaard might be willing to bite the bullet and insist that, contrary to appearances, deliberate immoral action *is* impossible and her failure to account for it is *not* an objection to her view. So I want to turn my attention now to what Korsgaard might say in response to my further contention that rendering deliberate immoral action impossible will result in an objectionable lack of respect for wrongdoers. Her response ultimately maintains that being responsible is not a fact about or characteristic of the agent but something we impose on him: we *hold* other people responsible.[88] But we only hold people responsible if we think they are the sort of people with whom we can enter into reasonable relationships. Thus, to hold someone responsible is to accord that person a basic kind of respect on her view, even if we only hold that person responsible for not exercising her capacity to rationally choose. I maintain that this respect is not sufficient to answer my concerns.

The idea that responsibility is a way of interpreting another's behavior rather than a property that the other person possesses follows straightforwardly from Korsgaard's commitment to the view that we interpret the world *as if* it were influenced by rational wills, even though as far as we know all our so-called 'actions' are physically

[88] "I argue that to hold someone responsible is to adopt an attitude towards him rather than to have a belief about him or about the conditions under which he acts." Christine M. Korsgaard, "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations" in *Creating the Kingdom of Ends* (New York: Cambridge UP, 1996): 188-221. See p. 188.

determined.[89] And it is precisely for this reason that holding others responsible is a mark of respect. If you always regarded others as though their behavior were determined, your relationships with them "would be, at best, like your relations to small children and the other animals."[90] Indeed, Korsgaard points out, "you cannot enter into *any* reciprocal relations with people whom you do not hold responsible."[91] Perhaps, even though we only hold people responsible for allowing themselves to be led by some impulse or another, the fact that holding others responsible also involves a judgment that the person as a whole is someone with whom we can still enter into relationships mitigates the charge that Korsgaard's account of responsibility leads to a problematic lack of respect for wrongdoers.

This suggestion seems even more plausible when considered in light of Korsgaard's account of different bad constitutions. After all, some constitutions, though defective, may be more unified than others or may resemble a virtuous constitution more closely. Korsgaard recognizes that in some circumstances it is tempting to "write somebody off,"[92] but she also observes that "[t]he extent to which we do this is a matter of degree."[93] In some cases, another person may be so irrational and disunified that relations with that person will be minimal at best, but in other cases the vicious agent

[89] This is the same as her commitment to the distinction between the theoretical and practical standpoints discussed in the next chapter. "Construed theoretically, responsibility is a characteristic of persons. Construed practically, holding one another responsible is something that we do, the more or less deliberate adoption of an attitude." Korsgaard, "Creating the Kingdom of Ends," 197.

[90] Korsgaard, "Creating the Kingdom of Ends," 207.

[91] Ibid.

[92] Ibid., 200.

[93] Ibid.

may resemble a unified agent enough that we can maintain fairly normal relationships with them.

Here's the problem though: even if it is a matter of degree how disunified a vicious agent is, they are *all* disunified. As a result, even if we can hold a vicious agent responsible because, as a whole, the degree of disunification is relatively minimal and we can thus still imagine maintaining relationships with him, it is nonetheless true that to some degree we 'write off' that person since insofar as he does wrong we do not think of him as an agent. In other words, my objection remains intact. Even if a person is unified enough to maintain normal relationships with her, insofar as I interpret her as acting wrongly, I interpret her as less than an agent and I must regard that action as comparable to the action of a child having a temper tantrum.

Ironically, among other things, in "Creating the Kingdom of Ends," Korsgaard is concerned to defend Kant from the charge that his standards of responsibility are too high. In particular, she worries that Kant cannot take into account mitigating circumstances that might excuse that person's behavior.[94] She even cites the passage where Kant suggests that respectful treatment of a vicious person should mirror respectful treatment of the person who makes logical errors. But Korsgaard claims that "regarding a person as stupid or making her errors seem reasonable are not our only options in these cases. Sometimes we can best preserve someone's self-respect, as well as our own respect for her, not by making her errors seem reasonable, but by laughing them off as the

[94] Again, this has to do with the distinction between the theoretical and practical standpoints discussed in the next chapter. Responsibility is the province of the practical standpoint, but mitigating circumstances tend to belong to the theoretical standpoint. Thus, it is hard to see how we could take mitigating circumstances into account when deciding whether to hold another person responsible.

result of transitory emotion or exhaustion. The same is surely true in the moral realm."[95] While I am willing to grant that this is sometimes true, the problem is that for Korsgaard we must always regard vice as the result of external forces like "transitory emotion" or "exhaustion." Again, surely it is objectionable to view all vicious actions this way, and Korsgaard's account of immoral action requires us to do so.

Nothing in Korsgaard's work on bad action or responsibility alleviates either of the concerns I have presented. Even if we were to overlook her failure to make sense of deliberate immoral action, the lack of respect towards those who have acted wrongly that seems to follow from her account remains a pressing problem for her project. Though we may accord wrongdoers a certain kind of respect by holding them responsible and thus showing that we regard them as capable of maintaining relationships with us, we must still regard the committing of the wrong as akin to a child throwing a temper tantrum. And this seems objectionable.

Two Objections Become One

Before proceeding to the next chapter, I want to tie together Cohen's critique of Korsgaard's earlier work with my concerns about her more recent work. I've already hinted at one connection between the two critiques: both critiques express concern that Korsgaard fails to adequately address certain kinds of character or actions exemplified in the person of the Idealized Mafioso. But the critiques each focus on a slightly different problem in Korsgaard's account. Cohen worries that Korsgaard fails to give an adequate account of morality that will rule out the actions and lifestyle of the Idealized Mafioso; I worry that, given the immorality of the Mafioso's actions, she fails to make room for the

[95] Korsgaard, "Creating the Kingdom of Ends," 211.

possibility of someone who deliberately chooses those actions. Even with this difference in emphasis, however, I think both critiques express a similar concern.

Namely, both critiques suggest that constructivism by its very nature *cannot* explain deliberate immoral action. Recall that Cohen described his objection as showing that Korsgaard's account was best understood as a description of "the experience or phenomenology of obligation, not its ground or authenticating source."[96] Korsgaard failed to provide an adequate explanation of why the Mafioso is *not* obligated – even forbidden – to act as he does because the Mafioso can *feel* obligated to act in a certain way even when he has no such obligation. Since constructivism attempts to draw normative conclusions from the rational processes of agents,[97] it is not at all surprising that its description of the ground of normativity would look like a description of the experience of obligation. Cohen accused Korsgaard of failing to make her account of normativity sufficiently guiding – that is, her account did not rule out enough objectionable and immoral behavior. If we assume that her account of normativity *is* sufficiently guiding, the fact that she attempts to draw these normative conclusions from the rational processes of agents means that immoral behavior cannot be the work of a rationally-functioning agent.

Both Cohen and I believe that an account of deliberate, immoral action requires us to appeal to some standard of normativity that is not the product of the rational processes of agents. In the next chapter I will look at Kant's response to similar concerns about his

[96] Cohen, 183.

[97] In *The Sources of Normativity*, it is the act of reflective endorsement that provides reasons and, by extension, the ground of obligation. In *Self-Constitution* it is the universalization of maxims that provides reasons and obligation – that is, the successful process of reasoning produces reasons. But in either case, as I say, normative conclusions are drawn from the rational processes of agents.

project. We will find that Kant attempts to appeal to just such a standard. Though it is not clear that Kant succeeds in resolving the challenges to his thought, I think it is telling that Kant recognizes that he needs a standard external to the agent's processes of reasoning. Unfortunately, though she claims to follow Kant, Korsgaard has effectively ruled out the possibility of making such an appeal herself.

CHAPTER TWO

Kant's Parallel Problem

Introduction

Korsgaard takes her project to be in keeping with Kant's work in ethics. Reflecting on the fact that her account of action seems to rule out the possibility of bad action, she remarks, "And of course there's an *exact* analogy to this difficulty in Kantian ethics. For a well-known problem in the *Groundwork* … is that Kant appears to say that only autonomous action, that is, action governed by the categorical imperative, is really free action, while bad or heteronomous 'action' is behavior *caused* by the work of desires and inclinations in us." [1] In this chapter I will strengthen my case for moral realism by considering Kant's response to this puzzle. I will show that Kant recognizes that making sense of free immoral action requires him to posit moral standards that are independent of the rational processes of deliberating agents – the sort of standards constructivists deny. Indeed, it is not even clear that Kant can make the maneuver he wants to precisely because of his own constructivist tendencies.

Kant scholars standardly regard his distinction between *Wille* and *Willkür* as providing his solution to the problem of immoral action, but Korsgaard appeals to a different distinction: the distinction between the theoretical and practical standpoints. I will first discuss Korsgaard's attempt to resolve Kant's problem. I will argue that her approach fails to appreciate the fact that the distinction between the theoretical and

[1] Korsgaard, *Self-Constitution*, 159-160, emphasis hers.

practical standpoints is the source of the problem. Thus the solution to that problem cannot lie in this distinction. Instead, I maintain that the distinction between *Wille* and *Willkür* is meant to bridge the gap between the theoretical and the practical standpoints. But as we will see, Korsgaard's constructivism prevents her from adopting this response.

Korsgaard's Interpretation of Kant's Solution

Though Korsgaard is well aware of the puzzle confronting Kant, in her work on Kant's ethics she says nothing about the distinction that is typically taken to be Kant's solution: the distinction between *Wille* and *Willkür*. *Wille* and *Willkür* are two different aspects of the will. *Wille* refers to the legislative capacity of the will and is closely identified with practical reason; *Willkür* is the executive capacity of the will. In other words, *Wille* ordains what ought to be done, but *Willkür* chooses how to act. There is some debate concerning how this distinction helps Kant.[2] Roughly, it is assumed that, though *Wille* exhibits autonomy in its legislation, free choice is exhibited by *Willkür* whether or not it adheres to the dictates of *Wille*. Though not an exact analogy, it might be helpful to think of the distinction between *Wille* and *Willkür* as similar to Augustine's divided will: just as Augustine describes himself as simultaneously desiring good but choosing evil, through *Willkür* the agent may choose to act wrongly even though *Wille* directs him to act rightly. Unfortunately, since Kant clearly maintains that autonomy is the 'positive concept'[3] of freedom, it is not at all clear that this distinction succeeds in

[2] Compare, for example, Lewis White Beck, *A Commentary on Kant's* Critique of Practical Reason, (Chicago: The University of Chicago Press, 1960) and Nelson Potter, Jr., "Does Kant Have Two Concepts of Freedom?" in *Akten des 4. Internationalen Kant-Kongresses*, ed. G. Funke and J. Kopper (Berlin: Walter de Gruyter, 1978), 590-596.

[3] See Kant, *G.*, 4:446-4:447.

helping Kant make sense of free immoral action.[4] After all, if free action is autonomous action, it seems that only action that conforms to the dictates of *Wille* is truly free.

Though Korsgaard does not explicitly appeal to the language of *Wille* and *Willkür*, in "Morality as Freedom" it is tempting to interpret her as making use of a similar distinction between the 'concept' of freedom and its positive 'conception.' Korsgaard adopts this distinction from Rawls, who describes a concept as merely formal while a conception fills in the substance of the formal concept. Alternatively, Korsgaard elsewhere describes a concept as referring to the solution to a problem, while a conception proposes a specific solution. The concept of justice, for example, simply refers to the solution to the problem of how we ought to distribute goods in society. Libertarians and socialists offer competing conceptions of justice by proposing specific solutions to the problem.[5] The basic idea is that a concept is a functional placeholder while a conception spells out what it is that performs that function.

For the purposes of our present discussion, the significance of this distinction is Korsgaard's application of it to Kant's account of free will. The function of free will is to cause action without being determined by outside forces. But this concept provides us with no information about the positive conception of free will, such as *how* it chooses without being so determined. Korsgaard explains that "the Kantian *concept* of free will would be 'a will which makes choices independently of all alien influences,' that is, a will that is negatively free. A positive *conception* of freedom would be a material

[4] Central to Potter's take on the *Wille*/*Willkür* distinction is the idea that Kant has one concept of freedom: "the *ability* of pure reason to be of itself practical." Kant, *MS*, 6:213. Freedom, according to this interpretation of Kant, is the capacity to be autonomous – not autonomy itself. I do think that Potter thus successfully defuses this problem. However, I think any account of *Wille* and *Willkür* struggles to explain how these two aspects of the *same* will, as I will discuss below.

[5] Christine M. Korsgaard, "Realism and Constructivism in Twentieth-Century Moral Philosophy," 302-326.

account of what such a will would in fact choose."[6] The concept of freedom negatively understood seems to correspond to the capacity for free choice associated with *Willkür* while the positive conception of freedom Korsgaard names – autonomous legislation – appears to correspond to the capacity of *Wille*.

However, just as appealing to *Willkür* to make sense of free immoral action remains problematic in light of Kant's identification of autonomy with the positive understanding of freedom, this distinction between a formal concept of freedom and its positive conception does not help Kant either. Because Korsgaard sees Kant as endorsing a single concept of freedom substantively filled out by the conception of autonomy, it still seems that the only action that is truly free is moral action. Nevertheless, Korsgaard maintains that this problem is illusory. She believes that Kant has *another* distinction in place that dissolves the apparent difficulty. More specifically, she believes that Kant can argue that free immoral action is not *theoretically* impossible. Scholars who think Kant's practical philosophy commits him to the position that we cannot hold people responsible for immoral action fail to take seriously the distinction Kant makes between the standpoints of practical and theoretical reason.

Korsgaard observes that in the *Critique of Practical Reason*, Kant claims that autonomy, the positive conception of freedom, is a postulate of practical reason.[7] She explains, "A postulate of practical reason is an object of rational belief, but the reasons for the belief are practical and moral… Although the beliefs are theoretical in form – the will is free, there is a God – their basis and function are practical. As Kant … constantly

[6] Christine M. Korsgaard, "Morality as Freedom" in *Creating the Kingdom of Ends* (New York: Cambridge UP, 1996), 159-187.

[7] See Kant, *KpV*, 5:132.

emphasizes in the second *Critique*, the postulates play no theoretical or explanatory role whatsoever."[8] Given that autonomy plays no theoretical role in Kant's philosophy, Korsgaard concludes that we cannot say with any certainty that moral evil *cannot* be freely undertaken. "A free but evil will," she writes, "is shown to be unintelligible *from the standpoint of pure practical reason*, but not to be theoretically impossible."[9]

The theoretical standpoint, briefly, is the standpoint from which we investigate. It is the point of view of a spectator. By contrast, the practical standpoint is the point of view of an agent. It is concerned with action, how I ought to interact with the world as it presents itself to me. This is why freedom is a postulate of *practical* reason: insofar as I am an agent I *must* regard myself as free and undetermined.[10] Korsgaard describes the difference between the theoretical and practical standpoints as follows: "The deliberating agent, employing reasons practically, views the world … as an expression of the wills of God and other rational agents… The theorizing spectator, on the other hand, views the world as … mechanistic and fully determined."[11] However, we can postulate that we are free to act morally because, though theoretical reason interprets the world as mechanistically determined, mechanism is a postulate of theoretical reason just as

---

[8] Korsgaard, "Morality as Freedom," 172.

[9] Ibid., 173, emphasis hers.

[10] Korsgaard asks the reader to imagine that he or she has a device implanted in the brain that will cause him or her to choose to do the opposite of whatever we would choose to do prior to the intervention of the device. (So imagine that, the instant you make a choice, the device activates and causes you to choose otherwise.) Insofar as you think of yourself as an agent, you do not want your choice to be determined by the device. However, so long as you are concerned with 'fooling' the device, you will be unable to act since you will constantly be worried that your so-called 'choice' is really determined by the device on your brain. In order for you to act, you must think of yourself as undetermined. See Korsgaard, "Morality as Freedom," 162-163.

[11] Korsgaard, "Morality as Freedom," 173.

freedom is a postulate of practical reason. Theoretical reason cannot rule out the possibility of freedom of the will.

The theoretical and practical standpoints map onto a distinction in Kant's metaphysics between phenomena and noumena. The phenomenal is the world of experience. Sensible 'intuitions,' which are something like raw sense data, are structured in accordance with the forms of intuition and the concepts of the understanding to produce experience.[12] Since we cannot experience anything apart from the structuring influence of the forms of intuition and the concepts of human understanding, phenomena are mere appearances. Something underlies these appearances, but we can never escape the influence of the concepts of the understanding to get at this 'something.' Thus, theoretical reason is restricted to dealing with the phenomenal realm. The 'something' underlying the phenomena, or, as Kant calls it, this 'thing-in-itself,' is the noumenal. We can admit that freedom is possible from a theoretical point of view since we cannot know the noumenal and thus cannot rule out the possibility of freedom.

Against Korsgaard's Interpretation

As we saw in the last chapter, Korsgaard does not appeal to the distinction between the theoretical and practical standpoints in dealing with her "analogous problem" about bad action. Instead, she proceeds to outline the 'bad constitutions.' I think Korsgaard is right not to invoke the distinction between the theoretical and practical standpoints in dealing with the challenge immoral action poses for her own thought and I think she was mistaken to invoke this distinction to resolve Kant's analogous problem.

[12] I will not discuss the forms of intuition or the concepts of the understanding. All that matters for my purposes is that these are structures imposed by the human mind onto sensible intuitions to produce experience.

In short, the distinction between the theoretical and practical standpoints cannot help Korsgaard because the critique of my first chapter is leveled from the standpoint of practical reason. In other words, the impossibility of deliberate immoral action in Korsgaard's thought is a problem *from the standpoint of practical reason*. Claiming that immoral action is *theoretically* possible, then, simply fails to address the issue.

The problem I pressed in chapter one is that we *do* interpret ourselves and others as choosing to act immorally, despite Korsgaard's claim that the process of deliberation that produces[13] genuine reasons for action leads us to conclude that it is impossible to do so. It is not simply that, having postulated the possibility of autonomous legislation, we have opened the door to the possibility that maybe, somehow, in some way someone *could* choose to act immorally. Rather, in interpreting the world as influenced by rational wills, we sometimes interpret them as choosing wrongly.

It may appear misleading to claim that this involves taking up the practical point of view since the practical point of view is so closely identified with the deliberative standpoint. But remember that the practical point of view interprets the world "as an expression of the wills of God and other rational agents."[14] Indeed, Korsgaard's account of responsibility depends on the fact that regarding others as rational agents requires us to adopt the practical standpoint. After all, she denies that responsibility is a trait people simply *have*; rather, we hold each other (and ourselves) responsible insofar as we take them to be rational agents. Korsgaard herself says, "Construed theoretically,

[13] I say "produces" to avoid suggesting that Korsgaard believes that we identify independently existing reason through the process of deliberation, but there is a sense in which genuine reasons are simply identified by this process. After all, each maxim is a candidate reason. It is only adopted as a genuine reason if it survives correct deliberative processes.

[14] Korsgaard, "Morality as Freedom," 173.

responsibility is a characteristic of persons. Construed practically, holding one another responsible is something that we do, the more or less deliberate adoption of an attitude."[15] Patrick Frierson remarks, "Korsgaard sometimes seems to associate the practical standpoint entirely with the *deliberative* or first-person perspective, although all that is strictly required is that one takes a standpoint according to which a person must be considered an *agent*, and this can occur whether one *deliberates* or is *evaluated* in a practical way."[16] Any time we interpret some feature of the world as an expression of a rational will, we take up the practical point of view. Thus, any time we take someone to be choosing to do wrong, we occupy the practical standpoint. And this is the problem facing Korsgaard: we can, from the practical point of view, interpret ourselves and others as deliberately acting wrongly.[17]

Kant's analogous problem also remains untouched by Korsgaard's solution in "Morality as Freedom." After all, if Korsgaard is correct in claiming that Kant maintains that immoral action is unintelligible from a practical point of view, then it is as much an objection to his view as to Korsgaard's that we so often interpret others and ourselves as choosing to act immorally.[18] I believe Korsgaard was mistaken to abandon the widely-accepted view that Kant intends to resolve the problem of free immoral action by

---

[15] Korsgaard, "Creating the Kingdom of Ends," 197.

[16] Patrick Frierson, "Two Standpoints and the Problem of Moral Anthropology" in *Kant's Moral Metaphysics*, ed. Benjamin J. Bruxvoort Lipscomb and James Krueger (New York: Walter de Gruyter, 2010), 83-110. See especially p. 85, emphasis his.

[17] Her response to the unintelligibility of immoral action from a practical point of view in "Morality as Freedom" is simply that immoral action *is* unintelligible. As I have already considered that response in chapter one, I will not readdress it here.

[18] If the problem Korsgaard addresses in "Morality as Freedom" is leveled from the theoretical point of view, then of course her solution applies to it. But even if that is the specific challenge to Kant's ethics that Korsgaard takes on in "Morality as Freedom," I think the analogue to her problem must be leveled from the practical point of view and is thus unaddressed by this article.

distinguishing between *Wille* and *Willkür*. In fact, I think Korsgaard's argument in "Morality as Freedom" moves in the opposite direction Kant moves in: rather than invoking the distinction between the theoretical and practical standpoints to deal with a problem unaddressed by the *Wille*/*Willkür* distinction, I think Kant intends the *Wille*/*Willkür* distinction to close the gap between the theoretical and the practical, the phenomenal and the noumenal.[19]

The Theoretical/Practical Distinction as the Source of the Problem

According to Korsgaard, the distinction between phenomena and noumena is "not between two kinds of beings, but between the beings of this world insofar as they are authentically active and the same beings insofar as we are passively receptive to them."[20] We cannot know 'noumena' because experience requires passivity: "we cannot experience activity as such."[21] However, she observes that we do regard ourselves as

[19] Korsgaard does not present her argument this way, of course, but given the ease with which her 'concept' and 'conception' of freedom map onto *Willkür* and *Wille*, respectively, this does seem to be a natural interpretation of her argument. However, the *Wille*/*Willkür* distinction is made explicit in Kant's writings after the distinction between the theoretical and practical standpoints had long been established, lending initial credence to my claim that this later distinction should solve problems unaddressed by the former, rather than vice versa.

[20] Korsgaard, "Creating the Kingdom of Ends," 203. Korsgaard thus clearly endorses what is called the 'two-standpoint' interpretation of Kant's metaphysics rather than the 'two-world' interpretation. According to the two-world interpretation, for any given item we find in the world, for example a chair, there are really two objects: the noumenal object and the phenomenal object. Correspondingly, there is a world of noumenal objects and a world of phenomenal objects, with the noumenal world standing in a causal relationship to the phenomenal world. Though different interpretations of Kant's metaphysics yield different interpretations of his ethics, I do not wish to adjudicate between 'two-standpoint' and 'two-world' readings of Kant's metaphysics. Not only am I interested primarily in Korsgaard's thought, but the problem I will press confronts both interpretations of Kant's metaphysics, so the fact that I will work only with the 'two-standpoint' interpretation should not matter.

[21] Korsgaard, "Creating the Kingdom of Ends," 204. Note that the distinction between phenomena and noumena was introduced as a result of our activity in structuring the intuitions that confront us. It is also worth pointing out that Henry Allison described the two standpoints in terms of considering things as they meet the epistemic conditions of the human mind and these same things considered "independently of the human mind and its cognitive apparatus." See Henry Allison, *Kant's Theory of Freedom* (New York: Cambridge UP, 1990), 4. I mention Allison's take on the two-standpoint interpretation because it is not

acting, even if we cannot experience "activity as such."[22] When we take up the perspective of an agent acting in the world, we must take ourselves to be acting independently of the laws of nature – that is, independently of the laws governing phenomena. In taking up the practical standpoint, then, we regard ourselves as noumena.

Of course, we do also regard ourselves as sometimes passive and thus as phenomena as well as noumena. And this is importantly related to the way in which the moral law binds us in the form of an imperative, or command. In order to act we must think of ourselves as setting ends for ourselves, as acting independently from the forces of nature, but at the same time we cannot help but admit that we are subject to various influences that might prevent us from achieving the ideal of autonomy. Autonomy, and by extension morality, are thus ideals we strive to conform to but do not automatically realize.

This dual perspective on ourselves poses several puzzles, especially when we must occupy the practical and theoretical standpoints at the same time. For example, deliberation typically requires the agent to take into account theoretical claims about the world, suggesting that in the act of deliberating the agent must somehow occupy two distinct standpoints. I think this puzzle – and others like it – can be resolved,[23] but there

---

obvious that his two standpoints map onto Korsgaard's two standpoints. But again, my concern is the relationship between Kant's thought and Korsgaard's.

[22] We must. Think again of Korsgaard's thought experiment discussed in n.9 above.

[23] See Frierson's "Two Standpoints and Moral Anthropology" for an overview of the challenges I have in mind as well as an attempt to address these challenges in light of a commitment to the 'two-standpoint' interpretation of Kant. A similar worry concerning the unity of the agent is especially vivid for 'two-world' theorists, since they take the agent literally to be a citizen of two distinct worlds. Jeanine Grenberg has recently suggested that the distinction between the theoretical and practical standpoints could be central to Kant's solution to this problem. See Jeanine Grenberg, "In Search of the Phenomenal Face of Freedom" in *Kant's Moral Metaphysics*, ed. Benjamin J. Bruxvoort Lipscomb and James Krueger (New York: Walter De Gruyter, 2010), 111-130.

is a related worry about immoral action that is similar to the problem I am pressing against Korsgaard. After all, immoral behavior supposedly results from a failure to act autonomously, allowing oneself to be determined instead by inclinations and other forces. So on the one hand, it seems that insofar as I take myself to be *acting* immorally, I adopt the practical standpoint. But insofar as I act *immorally*, I am determined to act as I do. If the practical point of view *can* take theoretical claims into account, then why can't the deliberating agent take into account the fact that she is determined to behave immorally? Does this account of the distinction between the practical and theoretical standpoints lead us to posit that an agent might simultaneously regard herself as both free *and* un-free from the practical standpoint? Insofar as she deliberates, she must consider herself free, but insofar as she takes into account the forces that determine her to behave immorally she thinks of herself as not free.

Frierson responds that, even if one regards oneself as under the influence of circumstances or psychology,

> One must still *decide* whether to give one's natural disposition the weight it typically has. One must decide whether these influences will have the causal power over oneself that they have been observed to have. And here, one *cannot* say, "well, it *has to* have that causal power over me," or rather, if one does say this, the "has to" will be a purely practical one, a decision about what one values, and not a decision "forced by the facts."[24]

I will return to whether this adequately addresses the problem for Kant, but first I want to consider whether this maneuver will work for Korsgaard. At first glance, this passage seems reminiscent of Korsgaard's description of responsibility as omission. Recall her claim that we blame you, not for the forces influencing you, but for the fact that you *let* them influence you. But I worry that Korsgaard's account of immoral 'action' in fact

[24] Frierson, 93-94, emphasis his.

undermines this response. After all, on Korsgaard's view, there can be no decision to act wrongly since there is no reason for the agent to act on.

To be fair to Korsgaard, she could say that the agent acts *as if* he had a reason when in fact he does not. But this does not solve the problem. In some ways, it magnifies it. Whether or not there *is* reason for action is a matter settled entirely from the practical standpoint. To say that he does not have a reason to act a certain way is to say that, from a practical standpoint, he does not *decide* to act in that way. Thus, when we regard someone as deliberately choosing to act immorally, we both regard him as deciding to act as he does and as *not* deciding to act as he does. The possibility of immoral action deepens the challenge of explaining the relationship between the theoretical and practical standpoints.

Insofar as we think of ourselves as agents, we take up the practical perspective, and we do sometimes think of ourselves and others as choosing to act wrongly. But at the same time, it is from the practical standpoint that we determine what reasons for action there *are*, and if there are no reasons for wrong action we *cannot* regard ourselves and others as choosing to act wrongly. In cases of immoral action, the practical point of view leads us to think of ourselves and others in terms more at home in the theoretical point of view. We can translate this problem into Kant's language: to occupy the practical standpoint is to think of ourselves as free, and Kant closely identifies freedom with the moral law. Thus, Frierson's response does not relieve all the pressure for Kant either.

Frierson may have shown that the deliberating agent cannot take the idea that various factors determine her action into account in her deliberations. However, he does

not explain how an agent who does wrong can reconcile her judgment that she freely chose to do wrong with her acknowledgement that, in doing wrong, she did not act autonomously and thus did not act freely.[25] Insofar as we have done wrong, we cannot think of ourselves as free. We are reduced to mere phenomena, acted upon by the forces of nature. In deciding to act wrongly – or in regarding ourselves (or others) as having acted wrongly – we seem to take conflicting perspectives simultaneously: we regard ourselves as both free and not free.

Notice that the problem is not that the standpoint of theoretical reason tells us we are not free while the standpoint of practical reason tells us we are; practical reason *all by itself* tells us that we are both free and not free. It is only from the standpoint of practical reason that we regard ourselves as agents, and as agents we must think of ourselves as free. At the same time, however, it is only from the standpoint of practical reason that the moral law applies to us as the complete expression of our freedom.[26] And this observation reveals something more: the problem *springs from* the way the distinction between the theoretical and practical standpoints is drawn. The practical standpoint is closely identified with both agency *and* morality, while immorality results from the causal forces at work in the theoretical standpoint. Thus, it is difficult to see how the practical standpoint *could* make sense of deliberate, immoral action. We need to be able to affirm *both* that we are agents *and* that, *qua* agents, we can (and do) sometimes act

[25] I do not intend to imply that Frierson was attempting to address this specific problem. I simply think Frierson has addressed a similar question and that his response might be mistakenly regarded as a solution to the problem I pose.

[26] Henry Allison argues that "moral requirements (or more generally, practical principles) must not only be addressed to agents conscious of having an intelligible character (in the language of the *Groundwork* a will) but also be addressed to them in that character." Allison, 226. 'Agents conscious of having an intelligible character' is equivalent to agents occupying the practical standpoint. Though I will not re-hash his argument, I am in agreement with him on this point.

immorally. The way that the distinction between the theoretical and practical standpoints is drawn, however, precludes us from doing so. Thus, the theoretical/practical distinction cannot be the *solution* to the problem.

The Wille/Willkür *Distinction Again*

Henry Allison suggests that the *Wille*/*Willkür* distinction is intended to solve just this sort of problem. He makes this suggestion while discussing the third section of the *Groundwork*. Kant's purpose in this section is to establish that there *is* a moral law which applies to human beings. Kant famously begins by trying to show that the mere fact that human beings must think of themselves as free – that is, as not determined by external forces – they are bound by the moral law. But this argument to establish the moral law is not enough. Remember that Kant believes the moral law is represented as an *imperative* for human beings, precisely because they do not always conform to the dictates of the law. Thus, Kant needs to show not merely that there *is* a moral law but that it can actually command beings that do not always conform to it. Kant puts the problem this way:

> All my actions as only a member of the world of understanding [the noumenal] would … conform perfectly with the principle of autonomy of the pure will [the moral law]; as only a part of the world of sense [the phenomenal] they would have to be taken to conform wholly to the natural law of desires and inclinations, hence to the heteronomy of nature.[27]

But there is more to it than that. It is a truism that Kant believes ought implies can. If there is an imperative that tells agents that they *ought* to act a certain way, they

[27] Kant, *G*., 4:453, bracketed text mine.

must be free to do so.[28] If they fail to do so, it cannot be because they were determined *not* to conform to the law. Thus, Kant's attempt to establish the categorical imperative is closely related to the issue that concerns me. Remember that I want to preserve the possibility of deliberate immoral action. Similarly, if Kant is correct in maintaining that the moral law takes the form of an imperative when applied to humans, he must preserve the possibility of free immoral action. Korsgaard has suggested that free immoral action *is* possible from the theoretical standpoint; it is only from the standpoint of practical reason that it is incoherent. However, I maintain that free immoral action must also be possible from the practical standpoint. Given that the categorical imperative is a vestige of practical reason – not theoretical reason – it seems that Kant would agree that free immoral action must present itself as a real possibility from the practical point of view. Kant's attempt to argue for the possibility of a categorical imperative, then, must address the fact that the practical standpoint appears to both require and rule out the possibility of free immoral action.

Unfortunately, the account Kant offers in the *Groundwork* is unsatisfactory. Kant seems to have failed to recognize the source of the problem since, rather than explaining how a view of the self that seems most compatible with the theoretical standpoint can have a home in the practical standpoint, he merely appeals to the distinction between phenomena and noumena as he has already drawn it. The distinction between phenomena and noumena, recall, maps onto the distinction between the theoretical and practical standpoints. Kant claims that, since the noumenal is more

[28] This is essentially Kant's point in *KpV*, 5:29-30. Some scholars believe Kant changes his argument in the second *Critique* to derive freedom from morality rather than morality from freedom. I take it that all that matters for my argument is that freedom and morality imply one another, and Kant seems to maintain his position in both the *Groundwork* and the second *Critique*.

fundamental than the phenomenal, even though we cognize ourselves as phenomena, when we act under the idea of freedom and thus as noumena, we recognize that we *ought* to conform to the moral law. To illustrate his view he asks the reader to imagine "the most hardened scoundrel" (so long as "he is otherwise accustomed to use reason").[29] Kant maintains that when confronted by examples of virtue, the 'scoundrel' will wish that he were similarly virtuous. He observes,

> He cannot indeed bring this about in himself, though only because of his inclinations and impulses; yet at the same time he wishes to be free from such inclinations, which are burdensome to him. Hence he proves, by this, that with a will free from impulses of sensibility he transfers himself in thought into an order of things altogether different from that of his desires in the field of sensibility, since from that wish he can expect no satisfaction of his desires and hence no condition that would satisfy any of his actual or otherwise imaginable inclinations.[30]

But this passage seems to reinforce the concern that Kant may not be able to make sense of immoral action from a practical point of view.

Notice, after all, that Kant's description of the scoundrel makes it sound as though he must think of his immoral actions as determined by his inclinations. If Kant means that the scoundrel must think of himself as determined from a *practical* standpoint, then the problem I pose remains intact. Remember that insofar as the scoundrel thinks of himself as *acting*, he must think of himself as *not* determined by natural forces. He *must* think of himself as free from the practical standpoint. Thus the practical standpoint, again, leads the scoundrel to conclude that he is both free and unfree. But insofar as the problem I articulated remains untouched, it is not clear how the moral law as a categorical imperative actually has any authority over the scoundrel's actions. Instead it

[29] Ibid., 4:454.

[30] Ibid.

looks like conformity to the categorical imperative is just a kind of vain wish of the scoundrel's. Wishing my behavior would conform to the categorical imperative when I consider my behavior determined is akin to wishing that I could fly: both involve a desire to be free from the laws that govern what I can and cannot do. Why should the fact that I wish that my determined behavior were different show that the categorical imperative has authority over me?

Alternatively, Kant could mean that the scoundrel must regard his actions as determined from the *theoretical* standpoint.[31] But in that case Kant is simply failing to address the problem of the agent who, in choosing to do wrong, is required by the practical standpoint to think of himself as both free and un-free. Insofar as he is not addressing such an agent, Kant is also not explaining the possibility of the categorical imperative, since the imperative addresses someone who recognizes his failure to conform to the moral law as a matter of concern from the standpoint of practical reason. Kant might be able to show that the moral law *does* have authority over the agent insofar as the agent considers himself solely as pure noumenon that always conforms to the law, but it is not clear how it could apply to him as someone who recognizes that he is also passively receptive to impulses and other natural forces. In other words: it does not explain how the moral law applies to someone who sometimes gives into those impulses. How could someone be bound by an imperative when we think of him as pushed around

[31] This passage provides a good example of how one's view of Kant's metaphysics influences one's reading of his ethics. The reading of this passage changes considerably if one endorses the 'two-world' interpretation of Kant's ethics rather than the 'two-standpoint' interpretation. On the 'two-world' interpretation, the scoundrel must be thinking of himself as determined as a member of the phenomenal world. Although the 'two-world' theorist will affirm that we take up the practical standpoint as agents, the 'two-world' theorist need not claim that in taking up the practical standpoint we automatically regard all our actions as noumenal. Nonetheless, I think the 'two-world' interpretation faces a similar problem: How does the free noumenal self have any influence on the determined phenomenal self? Or, rather, how could the determined phenomenal self have any obligation to adhere to the law of the noumenal self?

by various forces, like all the other parts of nature? Kant is simply not addressing the problem I posed – or the one he posed for himself.

Allison recognizes the problems with this argument for the possibility of the categorical imperative, observing that

> the fundamental problem is that the putative addressee of the law, the phenomenal self, that is, the agent qua considered part of the sensible world, is *ex hypothesi*, incapable of recognizing any obligation, much less of acting in light of the 'ought.' This is because beings so considered are not merely sensibly affected (which is the point Kant wishes to emphasize in explicating the imperatival form of the law) but causally necessitated.[32]

He continues, "With the benefit of hindsight, it seems clear that what Kant needs at this point but does not yet possess is precisely the *Wille-Willkür* distinction,"[33] a distinction that appears in Kant's later writings.

How would differentiating between *Wille* and *Willkür* help? Because the *Wille*/*Willkür* distinction is a distinction in the faculty of the will itself, both *Wille* and *Willkür* fall within the practical standpoint. In other words, we posit *Wille* and *Willkür* only when we adopt the practical point of view. The theoretical standpoint, which views everything as mechanistically determined, certainly would not posit a distinction in the will, since it would not posit a will that causes things independently of mechanistic forces at all. Since *Wille* legislates, but *Willkür* chooses, we can, from a practical standpoint, choose to do wrong even though from a practical standpoint we also legislate morality.

Allison puts the point this way: Kant's problem is that he is attempting to argue that the moral law is binding on beings that are considered causally necessitated from a theoretical standpoint, rather than on beings regarded from the practical standpoint as

[32] Allison, 226.

[33] Ibid.

having sensibly affected wills. It is the latter that Kant needs to argue. *Wille* legislates as pure practical reason, but *Willkür* identifies that aspect of the will that is sensibly affected. *Willkür* is, however, still an aspect of the will and thus posited by the practical standpoint. In other words, the distinction between *Wille* and *Willkür* makes room for agents to recognize themselves as influenced but not causally determined by incentives other than respect for the moral law. By appealing to this distinction, then, Kant can avoid having to claim that, from a practical point of view, agents who act immorally are both free and un-free.

Problems with the Wille/Willkür *Distinction*

If Korsgaard's distinction between the 'concept' of freedom and its positive 'conception' is meant to map onto the distinction between *Willkür* and *Wille*, then her tactic in "Morality as Freedom" is wrongheaded insofar as it attempts to appeal to the distinction between the theoretical and practical standpoints to solve a problem created by that very distinction. But she may not be wrong to avoid the distinction between *Wille* and *Willkür*. While the *Wille*/*Willkür* distinction purports to explain how someone can, from the practical standpoint, regard herself as choosing to act immorally, it is in fact a problematic distinction.

For one thing, it seems *ad hoc*. Apart from the fact that Kant's moral theory appears to preclude free immoral action unless he can invoke the distinction between *Wille* and *Willkür*, there doesn't seem to be any reason to posit a distinction between the will's legislative and executive capacities. Furthermore, Kant nowhere develops the idea of this distinction in the will. The idea that there are two distinct capacities that somehow

unite in a single faculty of the will is difficult to grasp, to say the least.[34] While I think we might be able to overlook the seemingly *ad hoc* nature of the distinction, I want to argue that, in light of the foregoing discussion, an elaboration of the relationship between *Wille* and *Willkür* is required. After all, if we cannot make sense of the two capacities uniting to form a single faculty of the will, this distinction will ultimately fail to help Kant make sense of immoral action.

If my conjectures above are correct, the *Wille*/*Willkür* distinction explains both the imperatival form of the moral law and our ability to act immorally since we no longer have to interpret the influence of outside impulses as the causally necessitating forces associated with the theoretical standpoint. Though these forces still influence our choices, they do not determine us to act one way or another. In the *Groundwork*, Kant claimed that the moral law took the form of an imperative when applied to human beings because we are affected by sensible impulses that can direct us to act wrongly. But at that point in his philosophical work it was not clear how these impulses could do anything but impel us as phenomena, creating problems for considering these impulses from a practical standpoint. It is this feature of sensible impulses that leads us to conclude that we are not free even as we regard ourselves as free. The distinction between *Wille* and *Willkür*, however, gives us the tools to regard sensible impulses as *influencing* but not determining us. We can reinterpret the role of sensible impulses in a

[34] Admittedly, not everyone finds these puzzles problematic. Gordon Michalson, for example, has suggested that "it is often the case with Kant's writing that a strained or awkward claim is best decoded, not with a view to resolving its tensions, but with a view to the dividends that it pays. In the case of the distinction between *Wille* and *Willkür*, the dividends are readily evident, especially if we view the distinction in the context of the problem of moral evil." Gordon E. Michalson, Jr., *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* (New York: Cambridge UP, 1990), 36.

way that is more comfortable within the practical standpoint. Sensible impulses no longer *impel* us; they merely affect us. We choose whether to act on them.

As promising as this maneuver sounds, however, the problem has not disappeared. The *Wille*/*Willkür* distinction enables us to regard sensible impulses as influencing agency – but not determining it – by identifying an aspect of the will that is affected by sensible impulses and another aspect of the will that legislates in accordance with the dictates of reason alone. It is generally accepted that Kant thinks only *Willkür* is affected by sensible impulses – not *Wille*, which he identifies with pure practical reason. But this is mysterious. After all, *Wille* and *Willkür* are two aspects of the *same* faculty. When someone chooses to do wrong, then, his will legislates that he do one thing and yet chooses to do another under the influence of his impulses. Given that his will orders him to do otherwise, how can he interpret himself as free in the instance in which he obeys his impulses rather than his will? How can he regard himself as deliberately *choosing* to do the wrong thing?[35] Once again we find ourselves in the position of treating the agent in a way that seems more compatible with the theoretical standpoint than the practical. After all, insofar as *Wille* is not subject to sensibility, it wills perfectly. Insofar as *Willkür is* subject to sensibility, it is not free (and insofar as it is not, it also wills perfectly).[36] If

[35] I certainly do not want to rule out the possibility that someone might experience a 'divided will' in which he "does what he does not want to do." See Romans 7:15-20. However, I am also concerned to retain the possibility of someone *deliberately* choosing to do wrong. This possibility has not yet been accounted for.

[36] On the 'two-world' interpretation sensible impulses automatically place the agent in the phenomenal world rather than the noumenal world. This does not mean that the agent cannot regard his impulses from a practical point of view, since for the 'two-world' theorist this means regarding the agent as a dual citizen of the phenomenal and noumenal worlds. But it does mean that the 'two-world' theorist has a similar problem insofar as she has to explain how the moral law that governs humans as noumena can bear on humans as phenomena who act in the world of experience. The *Wille*/*Willkür* distinction on a 'two-world' theory, then, might be regarded as an attempt to bring the phenomenal and noumenal aspects of humanity together in the will, but the interaction of these two aspects remains mysterious.

Kant does not have the resources to explain away this puzzle, we cannot simply proceed to use the *Wille*/*Willkür* distinction for the sake of the dividends it pays: its purpose within Kant's ethical system is undermined by the persistence of this puzzle.

Lessons from Kant for Korsgaard

I do not intend to argue that this distinction between an aspect of the will that legislates how one ought to act and an aspect that chooses how one will act makes sense or solves Kant's difficulties (in fact, as we will see shortly, I think this distinction points to a genuine tension in his thought), but I do think the way Kant has gone about trying to solve the puzzle about free immoral actions is revealing. What Kant has recognized is that, in order to make sense of free immoral actions, he must find a way to separate the *standards* of freedom from the free choice characteristic of the deliberation of agents. But this is precisely the sort of distinction Korsgaard cannot appeal to in order to explain the possibility of deliberate immoral action. After all, her constructivist project is to derive standards for action from the rational processes of agents.

According to Korsgaard, it is by following correct rational procedures that reasons for action are constructed, and it is only by acting on such reasons that we realize our agency.[37] To act on any incentive other than those that can be endorsed by properly engaging in rational reflection is to fail to be an agent. Indeed, I think recognizing that constructivists cannot appeal to this sort of distinction explains Kant's own problems in

[37] My critique in this chapter applies chiefly to Korsgaard's work in *Self-Constitution*. In *The Sources of Normativity* she seems to think that reasons are created by the act of reflective endorsement, whether or not the reflection was carried out correctly: "[I]t is the endorsement, not the explanations and arguments that provide the material for the endorsement, that does the normative work." Christine M. Korsgaard, "Reply" in *The Sources of Normativity*, ed. Onora O'Neill (New York: Cambridge UP, 1996), 219-258. See especially p. 257. I discuss reasons to be troubled by her account in *The Sources of Normativity* in my other chapters.

invoking the distinction, discussed above. Whether or not Kant is a constructivist is controversial, but I think it is fairly clear that constructivists are inspired by Kant's project of locating the binding force of the moral law in the legislation of the agent's own will.[38] Thus, he tries to place the legislative capacity in the will: *Wille*. But as a result it is not clear how the will could act contrary to those standards legislated by the will itself.

Korsgaard correctly assessed her own position when she noted that, on her view, "responsibility in general is going to look a lot more like responsibility for omission."[39] And as far as it goes, this view is consistent with much of what Kant said. However, Korsgaard does not fully appreciate the need for an account of deliberate immoral action or the fact that the impossibility of such an account poses a serious objection to her theory. Kant does recognize this need and attempts to find a way to locate the standards of freedom outside the free deliberations of agents, whether or not he succeeds in doing so. Unfortunately, Korsgaard has effectively ruled out the possibility of using similar tools to amend her theory.

Conclusion

It is not without reason that Kant scholars have traditionally looked on the *Wille*/*Willkür* distinction as the solution to the problem of immoral action in Kant's moral theory. Korsgaard's appeal to the distinction between the theoretical and practical standpoints to address the same problem mistakes the source of the problem for its

[38] I deliberately refer to the "legislation" of the agent's will to avoid presenting a constructivist reading of Kant rather than a realist reading of Kant. Patrick Kain, a staunch proponent of a realist reading of Kant, distinguishes between legislating the moral law and authoring the moral law. Human beings may legislate the moral law, but they are not its authors. See Kain.

[39] Korsgaard, *Self-Constitution*, 175.

solution. But her neglect of the *Wille*/*Willkür* distinction is a mark of consistency on her part. Korsgaard can appeal to no such distinction to make room for deliberate immoral action in her account of agency. If she is correct in maintaining that her problem is an analogue of Kant's, her inability to appeal to a distinction analogous to the *Wille*/*Willkür* distinction is unfortunate. As we have seen, this distinction, though undeveloped by Kant himself, reveals that Kant recognized what sort of solution to the problem of immoral action was needed – a solution constructivists cannot invoke.

CHAPTER THREE

Making Matters Worse

Introduction

The problem of immoral action is connected to the problem of moral skepticism. Korsgaard inadvertently rules out deliberate immoral action in her attempt to rule out moral skepticism. And we have seen that, understood in one way, the unintelligibility of immoral action renders Korsgaard's normative question nonsensical. Candace Vogler's project makes sense of immoral action but at the cost of being unable to supply a response to the moral skeptic. Indeed, Vogler doubts that any secular ethical theory *can* both make sense of immoral action and respond to the moral skeptic.[1] After all, if there is some reason to act immorally, how can we guarantee that someone will never have *most* reason to act immorally? Despite my contention that deliberate immoral action is impossible on her account, Korsgaard does suggest that it *is* possible – and even good – in certain nonideal circumstances. In this chapter I want to look at the special circumstances that make deliberate immoral action possible. I will argue that Korsgaard's contributions to nonideal theory lend credence to Vogler's argument that she must either fail to adequately account for immoral action or fail to respond to the moral skeptic.

Korsgaard acknowledges that we will sometimes be placed in circumstances in which moral action is impossible. In those cases, Korsgaard's attempt to constrain the moral skeptic by claiming that the structure of an action just *is* the structure of a moral action will no longer apply. Since moral action – and therefore action – is impossible,

[1] Though she thinks a theistic ethical theory can. See Vogler, esp. chapter 2.

why shouldn't the skeptic embrace immoral behavior? Korsgaard's only recourse in this situation is to appeal to the value of humanity. But as we have seen, there is reason to think this appeal is inadequate.

The distinction between ideal and nonideal theory, which belongs properly to political philosophy, is introduced in John Rawls's *A Theory of Justice*, where Rawls writes,

> The intuitive idea is to split the theory of justice into two parts. The first or ideal part assumes strict compliance and works out the principles that characterize a well-ordered society under favorable circumstances. It develops the conception of a perfectly just basic structure and the corresponding duties and obligations of persons under the fixed constraints of human life. My main concern is with this part of the theory. Nonideal theory, the second part, is worked out after an ideal conception of justice has been chosen; only then do the parties ask which principles to adopt under less happy conditions.[2]

I will not address debates about whether Rawls was correct to maintain that ideal theory should precede nonideal theory. What is important for my purposes is that ideal theory imagines a situation of 'strict compliance' – in which all members of the society adhere to the strictures of justice – and 'favorable circumstances,' while nonideal theory deals with 'less happy conditions.'

Rawls identifies two sources of nonideality: natural limitations and partial compliance, corresponding to the favorable conditions and strict compliance characteristic of ideal theory. I will focus primarily on nonideality that results from partial compliance. Rawls explains, "When we ask whether and under what circumstances unjust arrangements are to be tolerated, …[w]e must ascertain how the ideal conception of justice applies … to cases where rather than having to make adjustments to natural limitations, we are confronted with injustice. The discussion of

[2] John Rawls, *A Theory of Justice* (Cambridge, MA: Belknap Press, 1971), 245-246.

these problems belongs to the partial compliance part of nonideal theory."[3] In other words, I am concerned with nonideal theory that addresses how we should act when confronted by people who do not obey the norms of justice and, by extension, the norms of morality.

A Kantian Account of Nonideality

When Korsgaard first began dealing with the problem of nonideality, she did not seem to think, as she does now, that nonideal conditions might justify morally impermissible actions. Instead, she claimed that "morality itself sometimes allows or even requires us to do something that from an ideal perspective is wrong."[4] In other words, she attempted to provide what she calls a 'double-level theory' so that morality itself could account for the fact that nonideal circumstances seem to call for different actions than ideal circumstances. The difficulty with providing a satisfactory double-level theory, however, is that it is not clear whether, in providing a second level, you are not simply abandoning your original theoretical commitments.

Tamar Schapiro critiques Korsgaard's early work in nonideal theory for precisely this reason. She claims that Korsgaard's characterization of nonideal conditions as those in which "the special conception of justice cannot be realized effectively"[5] is problematic because it "makes it sound as if the essential problem is one of inefficacy" and "the existence of this sort of defect could not provide grounds for departing from ideal

[3] Ibid., 351.

[4] Korsgaard, "The Right to Lie," 135.

[5] Ibid., 148.

principles on a truly nonconsequentialist account of such principles."[6] Indeed, at one point Korsgaard explicitly claims that "[i]n evil circumstances, but only then, the Kingdom of Ends can become a goal rather than an ideal to live up to."[7] This remark does make it sound like Korsgaard's allegedly nonconsequentialist Kantian ethic ultimately rests on some conception of the good to be attained.

Schapiro's own project is to describe "what structure a nonconsquentialist moral theory has to have if it is to make sense of both purist and pragmatic intuitions with regard to the wrongdoing of others."[8] In other words, how could a nonconsequentialist make sense of the fact that the wrongdoing of others seems to affect how we should act? She proceeds by suggesting a nonconsequentialist account of 'blind alley' dilemmas, or dilemmas which force a virtuous agent to act with insufficient justification.[9] Essentially, Schapiro appeals to the notion of a practice, or communal activity. She believes that practices are governed by constitutive rules. Justification of actions within those practices must be given in terms of these rules. Why did Jimmy Rollins run from here to there? Because according to the rules of baseball, after you hit the ball you run from home plate to first base. Had Rollins run directly to second base, or started spinning in circles, or sung the alphabet song, we could offer no justification for his action that would make sense of it as a part of the game of baseball. But because practices are communal

---

[6] Tamar Schapiro, "Compliance, Complicity and the Nature of Nonideal Conditions," *The Journal of Philosophy* 100 (2003): 329-355. See pp. 331-332, n.4.

[7] Korsgaard, "The Right to Lie," 153.

[8] Schapiro, "Compliance, Complicity, and the Nature of Nonideal Conditions," 331.

[9] 'Blind alley' dilemmas are first discussed by Thomas Nagel and, according to him, are the result of a conflict between our allegiance to both consequentialist and nonconsequentialist considerations. See Thomas Nagel, "War and Massacre" in *Mortal Questions* (New York: Cambridge UP, 1979): 53-74.

activities, the noncompliance of others can make it so that no course of action available to you can be justified by the rules of the practice. You may find yourself in a 'blind alley.'

Schapiro suggests a helpful example: imagine that you enter into a negotiation with another party. Though you certainly intend to represent your own interests, you want to arrive at some sort of reasoned compromise with the other party. But if the other party, though nominally negotiating, is ultimately just stalling for time, there is a sense in which you are not negotiating. Your action, in this case, requires the compliance of the other party. As long as he refuses to seriously consider your point of view, you cannot negotiate; you can only 'babble.'[10] Schapiro summarizes: "[B]ecause you are a participant you have to play, and because the rules are constitutive of participation, the only way to play is to play by the rules. But because the background conditions presupposed by such rules are ill established, playing by the rules fails to amount to participation in the relevant form of activity."[11]

Furthermore, it seems that you have some reason to refuse to continue with this sham of a 'negotiation.' Indeed, many courses of action not normally open to someone 'playing by the rules' of a genuine negotiation may, under the circumstances, seem reasonable options. But why do other courses of action seem reasonable options? Remember, to be a nonconsequentialist account, it cannot simply be that the alleged negotiation's status as a sham fails to realize some end. Rather, the reason must appeal to the form of the practice itself. Schapiro's suggestion is that the decision to engage in a

[10] Tamar Schapiro, "Compliance, Complicity, and the Nature of Nonideal Conditions," 337. Interestingly, Schapiro also points out that the other party may prevent you from negotiating by supporting *your* interests. If he goes through the motions of 'negotiating' without seriously considering your arguments as part of the process of arriving at some kind of deal, the negotiation is a sham.

[11] Schapiro, "Compliance, Complicity, and the Nature of Nonideal Conditions," 340.

practice – a communal activity – is the decision to have your actions express a kind of shared will. Schapiro leaves the notion of a 'shared will' largely undeveloped, but she does tell the reader that she is assuming that the shared will is a kind of 'plural subject' that results from the web of claims the participants in the practice make on one another.[12] For now it is sufficient to note that the justification for departing from the rules of the practice is that noncompliance subverts the form of the practice as an expression of a shared will. Departing from the rules is *not* justified because noncompliance has made the practice in question ineffective in achieving some good.

Towards the end of her essay, Schapiro asks, "[W]hat if there were a non-optional practice, a role we had to occupy on pain of not being agents at all, and a set of rules we had to follow on pain of not doing anything at all?"[13] This question calls to mind Korsgaard's response to the skeptic in *Self-Constitution*, where morality *is* that set of rules we follow on pain of not doing anything at all. But does Korsgaard think of morality as a practice, a communal activity? The answer is unequivocally 'yes.' In "The Reasons We Can Share: An Attack on the Distinction between Agent-Relative and Agent-Neutral Values," she contends that "the primal scene of morality … is not one in which I do something to you or you do something to me, but one in which we do something together. The subject matter of morality is not what we should bring about, but how we should relate to each other."[14]

---

[12] Ibid., 343.

[13] Ibid., 355.

[14] Christine M. Korsgaard, "The Reasons We Can Share," 275.

Schapiro takes Korsgaard's more recent essay, "Taking the Law into Our Own Hands: Kant on the Right to Revolution," to offer a Kantian account of nonideal conditions consistent with her own. Korsgaard presents the article as addressing a specific question in political philosophy: Is revolution ever justified? Behind this question lies Kant's account of political society. He believes political society provides the condition necessary for rights and their enforcement.[15] Any attempt to enforce a claim of right implies a commitment to a system of rights existing between human beings. That is, I cannot claim to have a right without demanding that others respect it or recognizing that others have similar rights. The same is true for you and your right-claims. Thus, not only would there be no rights without political society,[16] but political society is taken to be the expression of 'a general will' for the mutual enforcement of rights.

This appeal to the 'general will' makes political society appear a candidate practice, in Schapiro's terms. Like Schapiro's shared will, this general will seems to be the product of a system of claims that people make on one another. But I don't think Korsgaard's account of the justification for revolt is consistent with Schapiro's account of the justification needed to depart from the rules of a given practice. According to Korsgaard,

---

[15] It is a bit misleading to say 'rights and their enforcement' since, for Kant, to have a right just *is* to have the authority to coercively enforce a claim. See Kant, *MS*, 6:231.

[16] Kant does think there is one innate right: "*Freedom* (independence from being constrained by another's choice), insofar as it can coexist with the freedom of every other in accordance with a universal law, is the only original right belonging to every man." Kant, *MS*, 6:237, italics in original. Indeed, it is this right that grounds the obligation to create political society. See Kant, *MS*, 6:312 and Kyla Ebels-Duggan, "Moral Community: Escaping the Ethical State of Nature," *Philosophers' Imprint* 9, no. 8 (August 2009), http://www.philosophersimprint.org/009008/ (accessed February 28, 2011). This fact, however, should not alter the subsequent account.

> [w]hen the very institutions whose purpose is to realize human rights are used to trample them, when justice is turned against itself, the virtue of justice will be turned against itself too. Concern for human rights leads the virtuous person to accept the authority of the law, but in such circumstances adherence to the law will lead her to support institutions that systematically violate human rights. The person with the virtue of justice, the lover of human rights, unable to turn to the actual laws for their enforcement, has nowhere else to turn. She may come to feel that there is nothing for it but for her to take human rights under her own protection, and so to take the law into her own hands.[17]

There is something about this passage that sounds tantalizingly close to Schapiro's account. For one thing, we have the makings of a blind alley dilemma: the virtuous person seems to be justified in both accepting and rejecting the authority of the current regime, making the justification for either course of action incomplete. Indeed, Korsgaard later describes this situation as one in which "morality ceases to give us clear guidance how to proceed."[18] But according to Korsgaard, it ceases to give clear guidance because it is not clear how best to respect rights, not because it is impossible to act in a way that accurately expresses the general will. Korsgaard's account still seems to appeal to consequentialist considerations.

What Korsgaard *should* have argued, to avoid appealing to consequentialist considerations, is that certain governments are not the expression of the general will. In that way her account might be made consistent with Schapiro's. But instead Korsgaard endorses Kant's own view that revolt is always wrong and all governments are legitimate.[19] For reasons which will become clear below, I think she must follow Kant on this point. However, I will first consider the problems that result from her insistence

[17] Christine M. Korsgaard, "Taking the Law into Our Own Hands: Kant on the Right to Revolution" in *The Constitution of Agency* (New York: Oxford UP, 2008), 233-262. See p. 257.

[18] Ibid., 260.

[19] Since Korsgaard presents the argument of this article as an attempt to resolve a tension in Kant's thought, it might be tempting to think that she does not necessarily endorse the view herself. However, as I hope the following discussion will make clear, I think there is reason to think that she does.

that revolt is always wrong before proceeding to explain why I think she is unable to claim that some governments are not legitimate because they do not reflect the general will.

The Wrongness of Revolution

Part of the reason Kant thinks revolt is never justified is that the act of the general will that grounds the institution of political society is a transcendental postulate of reason. In other words, it is something practical reason requires us to assume in order to explain the possibility of just interactions. It is not the case that some governments were produced by an act of the general will and some were not. Rather, when considering the present form of government, we must postulate that *it* expresses the general will of the people. And it does not matter how that government operates since the government's authority springs from its status as the expression of the general will of the people. The fact that you disagree with a particular law or judicial ruling does not entitle you to revolt because justice is determined by the general will and therefore by the government. Korsgaard explains, "If someone has enough authority to make and execute laws, and the people are living and acting and relating to one another under those laws, then that is their general will."[20] Revolution is therefore always proscribed.

Korsgaard suggests additional reasons to think that revolution is always, "in one fairly clear sense, wrong."[21] This remark is accompanied by a footnote in which Korsgaard muses, "What fairly clear sense? Not universalizable, certainly; but the more

[20] Korsgaard, "Taking the Law into Our Own Hands," 249.

[21] Ibid., 260.

important point is what that shows: that such an action relates us wrongly to others."[22] In her earlier work, Korsgaard claimed that a maxim was not universalizable if it created a practical contradiction in the agent's will. That is, if the purpose stipulated by the maxim would be thwarted by the universalization of the maxim, then it would not be permissible.[23] Korsgaard might think that even a seemingly good maxim directed towards revolution, like a maxim to revolt in order to respect human rights, cannot be universalized precisely because of the recognition that the government, as the expression of the general will for the mutual enforcement of rights, provides the necessary condition for the respecting of rights. Given this background, willing this maxim as universal law involves one in a practical contradiction. One cannot will to respect rights using means that undermine the condition of those rights. This certainly seems like a plausible interpretation of Korsgaard's position.

But in light of Korsgaard's later work, I think there is another reason why Korsgaard does not think the universalization test will ever permit revolution that also explains her further claim that revolution relates us wrongly to others. That is, in order for a reason to be universalizable it must be a public reason.[24] Public reasons appear

---

[22] Ibid., n.31.

[23] See Christine M. Korsgaard, "Kant's Formula of Universal Law" in *Creating the Kingdom of Ends* (New York: Cambridge UP, 1996), 77-105. Again, the innate right to external freedom would be the one exception, but I do not think this alters my account since it is the right to external freedom that requires us to enter into political society. The only way it could affect my account is if the reason external freedom requires us to enter into political society were unaddressed by that society. In that case, the alleged 'government' would not be a government at all. Since the problem addressed by political society remains unaddressed, we remain in the state of nature. As such, this 'government' could not be the expression of the general will. But that is precisely the alternative account I go on to consider.

[24] The shift from talking about maxims to talking about reasons is not unjustified. While Kant speaks frequently of maxims, Korsgaard talks about reasons for action. But our discussion of Korsgaard's take on reasons for action in chapter one should make it clear that Korsgaard takes herself to be talking about the same thing Kant refers to as a maxim.

early on in Korsgaard's work. In fact, it is by appealing to the publicity of reasons that Korsgaard argues in *Sources* that valuing my humanity – that is, valuing my reasons – commits me to valuing the reasons and humanity of others. Unfortunately, though Korsgaard's ethics depends on the publicity of reasons, she never clearly explains what it is for a reason to be public. I think we can get some sense of what she has in mind, however, by comparing her account of public reasons with the account of reasons offered by the Egoist. The Egoist is the person who insists that all of my reasons are strictly mine and all of your reasons are strictly yours. That you need help is not a reason *for me* to act in any particular way. I will only help you if *I* have a reason to help you – something that appeals to my own projects, goals, and desires. To conceive of reasons as public is to deny this claim. Hence, Korsgaard at one point describes public reasons as "reasons whose normative force can extend across the boundaries between people."[25]

Korsgaard thinks we experience the reasons of others as binding on us:

> If I call out your name, I make you stop in your tracks. (If you love me, I make you come running.) Now you cannot proceed as you did before. Oh, you can proceed, all right, but not just as you did before. For now if you walk on, you will be ignoring me and slighting me. It will probably be difficult for you, and you will have to muster a certain active resistance, a sense of rebellion. But why should you have to rebel against me? It is because I am a law to you. By calling out your name, I have obligated you. I have given you a reason to stop.[26]

Furthermore, she has recently suggested that personal interaction is possible only *because* reasons are public. Consider the act of promising. Korsgaard proposes that, even after making a promise, I am not bound until the promise is accepted. I could back out, or you could reject the promise, and I would not have failed to fulfill any obligation. But if, in effect, no promise has been made until the promise has been accepted, what is there to be

[25] Korsgaard, *Self-Constitution*, 191.

[26] Korsgaard, *The Sources of Normativity*, 140.

accepted? Korsgaard concludes that "promises … cannot be understood as the results of successive acts. Instead they must result from the formation of a single common will, from a moment of unity between us."[27] Something similar holds for all transactions. I cannot give you anything without relinquishing my claim to it, but then if you take it I have not given it to you: you simply found it in an unclaimed state. In order to actually transfer something to you, we must unify our wills. And this, in turn, requires us to think of reasons as public. We cannot unite our wills – that is, we cannot share in deliberation to arrive at a common conclusion about what to do[28] – unless our reasons transcend personal boundaries.

In *Self-Constitution*, Korsgaard explicitly connects the publicity of reasons to her interpretation of the universalizability requirement. She explains:

> If I conceive of reasons as private, and accept a universalizability requirement, I am committed to the view that if I have reason to do action-A in circumstances-C, then I must be able to grant that you also would have a reason to do action-A were you in circumstances-C… On the public conception of reasons, by contrast, a universalizability requirement commits me to the view that if I have a reason to do action-A in circumstances-C, I must be able to *will* that you should do action-A in circumstances-C, because your reasons are normative for me.[29]

Korsgaard's moral theory requires the latter conception of universalizability, not only because she thinks of morality as a shared project between persons, but because she thinks of *agency* as a project shared between my self at one time and my 'selves' at other times. This is how action binds the agent into a unified whole, by following from reasons that can be shared across time. It is not enough that right now I believe I have a reason to act in a certain way and am willing to grant that my future self, placed in similar

[27] Korsgaard, *Self-Constitution*, 190.

[28] Ibid.

[29] Ibid., 191, emphasis hers.

circumstances, would have reason to act in the same way. I have to be able will that *I* should act in this way, whether now or at a future time.[30]

We have now seen that *if* the reason for revolution is not a public reason it cannot be universalizable; but notice that the above quote also indicates that if a reason is not universalizable in a certain sense it cannot be a public reason. After all, we have only two options: either reasons are public or private. What is entailed by a universalizability requirement differs depending on which view of reasons you endorse. On the public conception of reasons, "a universalizability requirement commits me to the view that if I have a reason to do action-A in circumstances-C, I must be able to *will* that you should do action-A in circumstances-C."[31] If I cannot *will* that you should do action-A in circumstances-C, I am not acting on a public reason. And the revolutionary cannot will this. Indeed, Korsgaard admits as much, explaining that "[t]he revolutionary cannot claim he has a justification, in the sense of an account of his action that other people must accept."[32]

And this is why the revolutionary's action relates him wrongly to others. By undermining the ends of others – in this case, opposing the general will – for reasons those others need not share in,[33] the revolutionary acts paternalistically "when

[30] Korsgaard's position in *Self-Constitution* is foreshadowed in her response to Nagel in "Reply," pp. 222-229.

[31] Korsgaard, *Self-Constitution*, 191.

[32] Korsgaard, "Taking the Law into Our Own Hands," 258. Notice that Korsgaard does not simply say that the revolutionary cannot expect others to take on his cause but, further, that the revolutionary cannot expect others to accept the justification he offers for his cause.

[33] Sharing in someone else's reasons does not necessarily mean acting just as that person does. Korsgaard's discussion of 'ambitions' in "The Reasons We Can Share" is helpful here. In that article she suggests that sharing can come in degrees, ranging from simply respectfully understanding the other's actions and not preventing him from undertaking them to actively helping him achieve his ends to taking on

paternalism … is what he hates the most."[34] That is, he unilaterally imposes his will on others, despite their contrary aims and reasons. When you take morality to be a project shared between people, this kind of paternalism is especially troubling. Korsgaard acknowledges this as well, pointing out that she does not take the fact that the revolutionary's action relates him wrongly to others "as an incidental feature of wrong actions, a mere effect of the fact that the actions are wrong and therefore others don't want us to do them." [35] Rather, she takes "the way [an action] relates you to yourself and others as *of the essence* of the morality of an action."[36]

Justifying Revolution?

Nevertheless, Korsgaard still maintains that this does not rule out the possibility that the moral person might revolt: "The just person respects the rights of humanity, and for this reason respects the government that enforces those rights, and the juridical condition that makes their enforcement possible. But it is by no means obvious that a person who makes the rights of humanity his end would never, under any circumstances, oppose the extant government."[37] Korsgaard admits that this will be a difficult decision for the just person. After all, in revolting, she not only lacks the guidance of the principle of universalizability, but she risks replacing the current regime with a still more unjust regime, or creating an environment in which there is no justice at all for a time.

---

the very same ends for oneself. Insofar as the revolutionary cannot offer an account others must accept, he cannot expect them to 'share' his end in any degree.

[34] Korsgaard, "Taking the Law into Our Own Hands," 258.

[35] Ibid., 260, n.31.

[36] Ibid., 260-261, n.31, emphasis hers.

[37] Ibid., 256.

Nevertheless, insofar as the end of justice is the enforcement of rights and the system of justice has been turned against rights, the just person may come to believe that revolution is called for.

The consequences of this conclusion are more serious than Korsgaard lets on. For one thing, her work in *Self-Constitution* suggests that the revolutionary isn't really acting at all. Insofar as a maxim must be universalizable if it is to bind the agent into a unified whole, the revolutionary 'acts' in a way that does not constitute her agency and, to that extent, does not act. Furthermore, insofar as Korsgaard admits that the good person *can* revolt, even when the maxim behind revolution does not conform to the structure of a reason, we might wonder why, say, the Idealized Mafioso – or, if it turns out that the Idealized Mafioso *does* act on maxims conforming to the structure of a reason, the not-so-Idealized Mafioso – *cannot* depart from the structure of a reason.

I want to consider four possible responses to this worry. First, it might be thought that I am simply highlighting the tragedy of the revolutionary's situation. Perhaps, insofar as the virtue of justice is turned against itself, the revolutionary is in a situation in which no 'action'[38] is available to him. It is impossible to act as a unified agent. After all, if the lesson of Schapiro's noncompliant negotiator extends to the practice of morality, as Schapiro herself suggests, the noncompliance of others can make it impossible for us to act morally. I believe that this maneuver is unsatisfying.

Whether or not the revolutionary finds herself in a tragic situation, Korsgaard gives her some grounds for acting in a way that is not strictly moral. Korsgaard is committed to the following claims: (i) the revolt is morally wrong, (ii) there is no

[38] I will drop the scare quotes around 'act' and its derivatives from here on, but it should be understood that it is not clear whether the just revolutionary acts.

universalizable reason for revolution, but (iii) sometimes the good person, *qua* good, can revolt. What is built in to that 'can'? It certainly does not seem to be moral permissibility. I would suggest, instead, that the revolutionary 'can' revolt because there is a certain kind of justification for his action, even if it is not a justification that others must accept. Indeed, Korsgaard describes her project in "Taking the Law into Our Own Hands" as an attempt to describe a kind of case in which a good person decides that, "for moral reasons, she must take the law into her own hands."[39] This makes it sound as though the good person has *some* kind of justification for revolting.

Korsgaard explicitly denies only that the revolutionary's actions are justified in the sense that reasonable people *must* accept the revolutionary's reasons for them. But this isn't to say that reasonable people *cannot* accept these reasons. Presumably, the revolutionary herself is a reasonable person. Korsgaard's description of the revolutionary's position supports this interpretation since the revolutionary is forced to make a kind of judgment call. Korsgaard claims that imperfection in a system does not justify revolution, since revolution would always be in order if it did. The good person can only revolt when the system of justice has been perverted. But it is not clear when the imperfection of a system of justice becomes a perversion. Thus, in deciding to revolt, the revolutionary will not be able to justify her revolt to all reasonable people, but some reasonable people may accept or at least understand her judgment.[40]

[39] Korsgaard, "Taking the Law into Our Own Hands," 234. I am not sure how Korsgaard can say this, given her view about the nature of reasons.

[40] Notice how this description corresponds to what universalization looks like when applied to private reasons. The revolutionary can grant that anyone in circumstances-C has a reason to do action-A, but the revolutionary cannot take the stronger position and will that anyone in circumstances-C do action-A.

Notice, however, that the revolutionary lacks the same justification the person who acts immorally lacks. According to Korsgaard's work in *Self-Constitution*, we are supposed to be constrained to act morally precisely because we can *only* act on universalizable maxims.[41] Observing that the revolutionary may be in the unfortunate position in which no *true* action is possible does nothing to change the fact that, in these circumstances, there is some sort of justification for behavior that is not strictly ethical. There doesn't seem to be any reason to assume that action we would more readily deem wrong could not appeal to some alternative justification as well. But it might be objected that this is no serious threat to Korsgaard's response to the skeptic since recourse to alternative justification is confined to a specific sort of situation: a situation in which morality offers no clear guidance, a 'blind alley' dilemma.

I have a two-part response to this maneuver. First, I think these 'blind alley' dilemmas are potentially far more frequent than this argument acknowledges. Korsgaard herself extends her account of the justification for revolution to other cases: "revolution is only one case, the most vivid perhaps, in which good people take the law into their own hands. Another, much more common and familiar sort of case, is when we paternalize an adult human being who is engaged in some sort of self-destructive behavior."[42] After all, "[t]he structure of the problem we face in these cases is exactly the same as that of the problem faced by the revolutionary. When we see someone perverting or destroying the humanity or autonomy in his own person, our respect for his humanity or autonomy is

[41] Again, where the account of universalization depends on a public conception of reasons.

[42] Ibid., 258.

turned against itself."[43] And this is troubling. Whenever we find ourselves in situations in which no course of action is fully justified, we will not be *constrained* to act in any way in particular. The frequency with which we find ourselves in this position should be alarming to anyone attempting to respond to moral skepticism.

Furthermore, notice that this once again calls attention to the need for a substantive account of what it means to value humanity. As I pointed out in chapter one, the call to 'value humanity' needs to be filled out more. Interestingly, Korsgaard herself suggests that most of us would intervene to save someone who was suicidal, "at least unless the person was hopelessly ill and in great pain."[44] Clearly, then, she doesn't take 'valuing humanity' to consist in saving humanity at all costs, but nor does it consist in respecting the reasons of another if those reasons conflict with whatever it means to 'value humanity.' Korsgaard offers no guidance to the agent who is trying to decide whether or not to intervene in another person's life. Of course, that is precisely Korsgaard's point when she claims that morality ceases to give guidance in some cases, but it is more problematic than she realizes. Unfortunately, certain substantive accounts of what it means to value humanity may result in actions that she would not want to regard as in any way justified, as in the case of the Idealized Mafioso.

Another response to my concern might seek to place constraints on how we can act in the face of immorality. Even if these 'blind alleys' occur more frequently than some might wish to admit, they do not supply the agent with an excuse to act in any old

[43] Ibid.

[44] Ibid.

way he wants. Schapiro, again, appears helpful here.[45] In a more recent work, Schapiro focuses on Kant's strident demand for honesty in all circumstances. But Schapiro has a rich conception of honesty in which honesty requires an acknowledgement of the other person's ability to govern herself. As a result, Schapiro concludes that "honesty can in principle come to be hollow when those to whom the truth is told are already held hostage to nonrational forces. Honesty is in these cases reduced to mere truth telling."[46] In such cases, the good person should try to act in a way that best approximates the ideal of honesty, even if that means deceiving the other person. However, this deception must "be undertaken in the spirit of honesty,"[47] with the good person taking care "to avoid interference with autonomy to the extent that it does exist and to foster or at least not preclude the development of autonomy where it does not yet exist."[48]

I have two concerns about this position. First, I do not think Schapiro successfully rules out the Idealized Mafioso. Of course, Schapiro acknowledges that the position sketched above is not enough. We can imagine that the murderer at the door and the suicidal person are in full possession of their rational faculties and deliberately choose to act as they do.[49] Schapiro concludes that "if there is anything about the murderer's

---

[45] Though she certainly does not speak for Korsgaard, she clearly shares some key convictions with her and it would not be implausible to think that Schapiro's ideas might be incorporated into a kind of Korsgaardian constructivism.

[46] Tamar Schaprio, "Kantian Rigorism and Mitigating Circumstances," *Ethics* 117 (2006): 32-57. See p. 47.

[47] Ibid., 48.

[48] Ibid.

[49] Korsgaard acknowledges this possibility with the suicidal person. See Korsgaard, "Taking the Law into Our Own Hands," 258. But this is a case in which she thinks morality clearly ceases to give guidance. It is less clear what she would say about the murderer at the door, especially in light of her work in *Self-Constitution*.

condition or orientation that could serve to mitigate the prohibition on deception, it would have to have something to do with the fact that the murderer is a bad guy, that he has adopted and plans to pursue a blatantly immoral … end."[50] Being a Kantian, Schapiro must somehow connect the badness of the murderer to his rational capacities, and she does so by maintaining that the murderer acts on reasons that cannot be shared. That is, the murderer refuses to act as a member of the Kingdom of Ends.

Perhaps in the case where the murderer pursues a victim for no recognizable purpose Schapiro has a point. But what about the Idealized Mafioso? Could we not imagine a Mafioso who does not simply acknowledge that others would have a reason to act as he does in certain circumstances, but who universalizes his maxim in accordance with a *public* conception of reasons? Why could he not kill or maim someone insufficiently loyal to the family *and* will that anyone else in those circumstances do the same thing? Or consider a less fanciful example, like Korsgaard's deliberate suicide. Korsgaard and Schapiro need a more substantive account of what can be morally willed.

But second, and more importantly, Schapiro's position simply fails to address the issue I think 'blind alley' dilemmas present. Though it might look like she places limits on what behavior would be acceptable in such situations, her claims will not compel the normative skeptic. After all, Korsgaard wants to say that the skeptic must act morally on pain of not acting at all. To fail to act morally would be to fail to act for any reasons at all. In nonideal circumstances, the agent *cannot* act in Korsgaard's sense.[51] Schapiro might recommend that the agent do his best to realize the ideal, but there is no reason that

[50] Schapiro, "Kantian Rigorism and Mitigating Circumstances," 51.

[51] Or, as in the case with the murderer at the door, though the agent can act, it doesn't seem like the full-blooded action available to the agent – telling the truth – is really what the agent should do.

will constrain him to do so. What can Korsgaard or Schapiro say to the moral skeptic in this case to demonstrate that he *must* act, say, "in the spirit of honesty"?[52] Given that action is impossible, that no maxim conforms to the structure of a practical reason, practical reason cannot require the skeptic to approximate action.

Before going on, it might be worth pointing out that I have difficulty imagining what other recourse a Kantian constructivist might have. On the one hand, Korsgaard sets the limits of morality by appealing to autonomy; on the other, Schapiro sets those same limits by appealing to the shareability of reasons.[53] If reasons must be either private or public, and if neo-Kantians must set the limits of morality by appealing to the demands of reason, we may be justified in thinking that these attempts exhaust the neo-Kantian arsenal's possible responses to the problems raised by nonideal theory.

Two Kinds of Skepticism

Finally, a more plausible response to the critique I've presented might draw on a distinction Korsgaard makes in *The Sources of Normativity* between the complete normative skeptic and the moral skeptic. It might be thought that my argument only undermines a response to the complete normative skeptic, not to the moral skeptic. Korsgaard openly acknowledges that her work in *Sources* will not convince the complete normative skeptic he must act in certain ways. In a discussion of suicide at the end of *Sources*, she admits that there is nothing she can say to the person who denies the value of her own life, explaining that

[52] Schapiro, "Kantian Rigorism and Mitigating Circumstances," 48.

[53] This is not to deny that Korsgaard believes reasons are shareable. Clearly she does. Indeed, it seems that her constructivism requires her to maintain this position. But I believe an appeal to autonomy that relied on a private conception of reasons would be vulnerable to many of the same critiques as Korsgaard's appeal to autonomy, and perhaps more besides. Furthermore, I think Schapiro attempts to capitalize on the shareability of reasons in a way that Korsgaard does not.

> [t]he price of denying that humanity is of value is complete practical normative scepticism. The argument is, if successful, a reply to the *moral* skeptic, one who thinks that he can value *something* without acknowledging the force of moral obligation. But I have not shown that *complete* practical normative scepticism is impossible. Is there an argument against that kind of scepticism, a reason not to commit suicide?[54]

At the time she wrote *Sources*, Korsgaard thought not – or at least that there could be no philosophical argument against that kind of skepticism. But her most recent work in *Self-Constitution* may be taken as an attempt to provide this response to complete normative skepticism by grounding normativity in the inescapability of choice and action rather than the contingent valuing of humanity. Since this maneuver is what causes the problem in chapter one that has been the continued focus of this present chapter, perhaps we must only reject her response to the complete practical skeptic.

I am open to this critique. It might be thought that the fact that the revolutionary's basis for action is not a public reason might pose a problem for Korsgaard's response to the moral skeptic in addition to her later response to the complete practical skeptic. After all, if Korsgaard is willing to admit that the just revolutionary has reasons for revolting, as she claims, she seems committed to the idea that it is possible to have reasons which are *not* public.[55] But Korsgaard depended on the claim that all reasons are essentially public to explain how valuing the humanity of my own person commits me to valuing the humanity of others. Indeed, if reasons are public, as she claims, there is nothing to explain. If it is not the case that all reasons are essentially public, this move seems suspect. But I think it might be possible to defend

---

[54] Korsgaard, *The Sources of Normativity*, 163, emphasis hers.

[55] Consider the following passage: "If one's end cannot be shared, and so cannot be an object of the faculty of desire for everyone, it cannot be good, and the action cannot be rational." Christine M. Korsgaard, "Kant's Formula of Humanity," 116.

Korsgaard's position by arguing that it is only in very special circumstances that the publicity of reasons breaks down. Such a maneuver would have to give an account of the publicity of reasons that explains why, in these circumstances, it evaporates.

In *The Sources of Normativity*, Korsgaard observes that there are two ways to explain the public nature of reasons: one could either ground publicity in objective values that exist out there in the world we share or one could claim that reasons are simply "inherently shareable."[56] This idea – that reasons might be inherently shareable without appealing to anything out there in the world – is opaque. Korsgaard's discussion of intersubjectivism in "The Reasons We Can Share" provides the best sketch of what she has in mind. In that essay, she contrasts her public conception of reasons with the view of reasons proposed by the Objective Realist.[57]

"The Reasons We Can Share" is a response to Thomas Nagel's argument in *The Possibility of Altruism* that attempts to establish that "an individual may have agent-relative or subjective reasons which have a legitimate normative force for her, but which have no normative force for others."[58] To support this claim, Nagel offers examples of activities a person might undertake, such as climbing Kilimanjaro, that do not seem to offer any agent-neutral reasons for action. That is, such activities presumably do not aim

[56] Korsgaard, *The Sources of Normativity*, 135.

[57] For our purposes I want to imagine that the Objective Realist rejects the Egoist's account of reasons in order to bring out a different contrast with Korsgaard's account of reasons than the contrast brought out by comparing Korsgaard's view of reasons with the Egoist's.

[58] Korsgaard, "The Reasons We Can Share," 282.

at any objective value that would demand certain actions of any person. Not everyone has a reason to climb Kilimanjaro. Korsgaard calls such projects 'ambitions.'[59]

At first glance, ambitions present special problems for Korsgaard's intersubjectivist position. After all, Korsgaard is committed to saying that your reasons are inherently shareable, but it is not clear what claim *your* idiosyncratic projects could have on *me*. Nagel, though a realist about certain kinds of value, believes such idiosyncratic projects could only make claims on those in close personal relationships with you. But Korsgaard points out that people with ambitions are not driven by thoughtless impulses. Rather, they can usually speak quite eloquently about *why* they want to climb Kilimanjaro, or write a best-selling novel, or amass a collection of rare birdhouses. The reasons they supply are not objective reasons waiting to be discovered out there in the world since, as pointed out above, not everyone will have a reason to climb Kilimanjaro or collect birdhouses. What we witness when we listen to people talk about their ambitions is the ability of human beings to project value onto the world through a process of reflective endorsement. And this demands our respect: your reasons for collecting rare birdhouses may not count as reasons for *me* to start collecting birdhouses, but, upon hearing you explain your hobby, I may no longer find it absurd, or I may alert you to unusual-looking birdhouses I've come across in case they hold some interest for you, and thus exhibit respect for your enterprise.

Korsgaard thinks intersubjectivism is uniquely positioned to explain *both* the fact that ambitions do not seem to appeal to any objective values *and* the fact that people are nonetheless able to explain these projects in ways that demand respect from others, no

[59] Nagel claims that deontological reasons are agent-relative as well, and Korsgaard proceeds to explain how these reasons are also intersubjective. For my purposes a discussion of ambitions should suffice.

matter how idiosyncratic the project. It is at this point that the contrast between Korsgaard's account of public reasons and the account offered by the Objective Realist becomes clear. Korsgaard writes, "The Intersubjectivist sees the other as human, and *therefore* shares or tries to share the other's ends. That is why she helps others to pursue their ambitions. But the Objective Realist sees no reason to help unless he *first* sees the other's ends as ones that he can share."[60] In other words, respect for humanity is the basis for the sharing of reasons.

If this interpretation of Korsgaard is correct and the publicity of reasons itself follows from our already deep-seated commitment to the value of others, Korsgaard may be right to maintain that the valuing of autonomy turned against itself can justify paternalism, where paternalism involves rejecting the publicity of reasons. After all, in the kinds of situations Korsgaard describes, valuing the humanity of another person urges us both to share her reasons and to override her reasons.

This might seem problematic for Korsgaard since, on this account, the publicity of reasons depends on a prior commitment to the value of others and, by extension, to morality. One might wonder, then, how the response Korsgaard offers in *The Sources of Normativity* is any better than the response offered by the realist. Remember, Korsgaard complains in *Sources* that the realist does not actually offer a response to the skeptic, but merely digs in his heels and insists that *x* is valuable and *y* is normative. It now looks like Korsgaard's response to the moral skeptic is: you ought to value humanity because you

[60] Korsgaard, "The Reasons We Can Share," 290, emphasis hers. It is not clear that Korsgaard represents the realist fairly here. It may not be that the Objective Realist first sees someone as pursuing ends and then simply decides whether or not to take those ends as ones he can share. Rather, it may be that insofar as the Objective Realist sees the other as acting *for ends*, he must see the other's ends as ones he can share. See Talbot Brewer's discussion of Elizabeth Ancombe's famous example of the desire for a saucer of mud in Talbot Brewer, *The Retrieval of Ethics* (New York: Oxford UP, 2009), 22-23. Cf. G. E. M. Anscombe, *Intention* (1957; repr., Cambridge, MA: Harvard UP, 2000), 70-72.

do. But isn't the moral skeptic asking why she *should* value humanity, particularly the humanity of others?

Korsgaard contends that it is okay to appeal to what human beings are like – particularly to our social nature – if that nature is 'deep.'[61] She seems to mean something like 'if that nature is inescapable.' She goes on to explain: "So the kind of argument we need here is … one that acknowledges that our reasons were never more than incidentally private in the first place."[62] The publicity of reasons is supposed to run so deep in human nature that the apparent privacy of reasons is ephemeral, accidental. Korsgaard compares passing over someone else's reasons to hearing the words of a language you know as mere noise. It is not clear how the argument she proceeds to offer works,[63] but her observations about human interaction cited above may prove compelling in their own right.[64]

In the next chapter I will reconsider whether this maneuver significantly distinguishes Korsgaard's response to the moral skeptic from responses offered by moral realists. For now we need merely note that Korsgaard herself seems to think her argument that 'we ought to value humanity because we do' will only work if she can show that reasons are practically inescapably public. That is, it has to be the case that, in normal circumstances, we cannot *help* but regard one another's reasons as making claims

---

[61] Korsgaard, *The Sources of Normativity*, 136.

[62] Ibid.

[63] The argument attempts to apply Wittgenstein's private language argument to reasons. Though I do not undertake to explore this argument here, it is worth pointing out that Korsgaard, though highly suggestive, never explicitly states how the analogous argument pertaining to reasons works. As it is, the proper interpretation of Wittgenstein's original argument is controversial. Spelling out Korsgaard's analogy would require more space than I have.

[64] I have in mind her reflections on what happens when I call out your name and the necessity of publicity to make sense of our interaction with one another.

on us. But it should not be surprising that in some circumstances it is not clear whether valuing another person leads us to share her reasons or override them.

My argument that those circumstances arise fairly frequently will not matter as a response to Korsgaard's work in *Sources* as it did in addressing Korsgaard's more recent work in *Self-Constitution*. The problem there was that there was no constraint to act one way rather than another when confronted with a blind alley dilemma, since no course of action conformed to the structure of a practical reason. Given the frequency with which blind alley dilemmas arise, Korsgaard should be uneasy with this conclusion. In this case, however, assuming Korsgaard has successfully argued that insofar as you value anything you value humanity, some constraint does remain in place: whatever you do, you must value humanity.

However, though Korsgaard's response to the moral skeptic in *Sources* may not be completely undermined by her work in nonideal theory, it does not remain unscathed. Though this point has probably become a bit repetitive by now, I think it is worth pointing out that attempts to rescue Korsgaard's theory from my critique have only made the need for a more substantive account of what it means to value humanity more apparent. The Idealized Mafioso remains a threat. Not only does Cohen's critique of Korsgaard's project in *Sources* remain largely in force, but a response to it becomes more urgent. Again, it seems that Vogler is right: Korsgaard can only rule normative skepticism out entirely by making immoral action impossible. Her previous response to moral skepticism – even though specifically targeting moral rather than complete normative skepticism – was inadequate precisely because it failed to sufficiently rule out immoral action.

Denying the Legitimacy of Unjust Governments

But perhaps Korsgaard could have avoided all this trouble simply by arguing that in fact there *are* cases in which the good person *is* fully justified in revolting. As I suggested above, I think Korsgaard would have to maintain that, in these cases, the government is *not* in fact the expression of the general will. But how could she maintain this position if the legitimate authority of the government is a transcendental postulate? To make this move, I think she would have to argue that it would be *impossible* for the general will to give rise to such a government. I want to very quickly sketch why I think she cannot make this argument.

Basically, I do not think it is clear *how* Korsgaard could argue that any given government could not possibly express the general will. My best attempt at filling in the argument for her would appeal to what the general will is supposed to be. Remember, this general will is directed towards a specific end: the mutual enforcement of rights that do not exist apart from this general will. Without this general will, we remain in a state of nature in which each of us has only the innate right to freedom "from being constrained by another's choice."[65] Not surprisingly, such a state is characterized by instability and violence. To actually succeed in acting freely and pursuing our chosen ends, we need to enter into civil society, where further rights can be established and enforced.[66] In other words, where there is civil society – or the expression of a general will to the mutual enforcement of rights – there is order.

[65] Kant, *MS*, 6:237. Kant adds the qualification: "insofar as it can coexist with the freedom of every other in accordance with a universal law." Ibid.

[66] Again, see Kant, *MS*, 6:312 and Ebels-Duggan, "Moral Community," especially pp. 2-8.

But this will not help Korsgaard. The only governments ruled out as illegitimate are those that are so poorly run that their society is disordered and anarchic. Remember, Korsgaard herself noted, "If someone has enough authority to make and execute laws, and the people are living and acting and relating to one another under those laws, then that is their general will."[67] Unfortunately, Korsgaard clearly wants to make room for the good person to revolt against well-ordered governments he deems unjust, despite the fact that those governments provide the very condition for justice.

Alternatively, it might be thought that she could try a maneuver analogous to one she makes in explicating the Formula of Humanity. Since she needs to argue that the government could not possibly be the expression of the general will, her claim that the Formula of Humanity rules out actions to which others could not possibly consent might sound like a promising starting point. Unfortunately, it is not clear to me how this argument might be extended to Kant's political theory. It might give us reason to think that governments that use methods ruled out by the Formula of Humanity understood in this way are acting wrongly, but this gives us reason to reform the government – not to revolt and undermine the condition of justice.[68] Besides, all governments will probably be characterized by these types of actions to some extent. As Korsgaard says, the mere presence of such actions cannot justify revolt. The presence of such actions would have to indicate that justice had been perverted, but there are no clear criteria for determining when that has happened. But these reflections may all be beside the point: as already mentioned in chapter one, Japa Pallikkathayil has raised serious problems for this reading

[67] Ibid., 249.

[68] See Kant, *MS*, 6:321-322. See Pallikkathayil.

of the Formula of Humanity, making it even more unlikely that it could provide us with some way to determine when a government is legitimately expressing the general will.

I see no other way that Korsgaard could maintain that it would be impossible for a given government to express the general will. And since we must postulate that governments are the expression of the general will, I do not know how she could maintain that any given government is *not* the expression of the general will without showing it to be impossible. But if we must regard a given government as the expression of the general will, then both Kant and Korsgaard must regard revolution as wrong and my foregoing critique of Korsgaard's work in nonideal theory remains in force.

Conclusion

Korsgaard's willingness to admit that in certain nonideal circumstances there is some justification for acting in a way that is, strictly speaking, wrong poses problems for her project of responding to the normative skeptic. There is reason to think that such circumstances arise quite frequently, especially in light of her reluctance to offer a more substantive account of what it means to 'value humanity.' And when these circumstances arise, the form of an action no longer has any constraining power since no 'action' is possible in such circumstances. If she is to constrain the behavior of agents placed in such unfortunate circumstances, I believe she will have to appeal once again to the demand that we 'value humanity.' But some immoral behavior can be described as expressing value for humanity. Vogler's dilemma reappears: Korsgaard could reject her work in nonideal theory and thus maintain the position articulated in *Self-Constitution*, which responds to normative skepticism by making deliberate immoral action impossible,

or she could continue to endorse her work in nonideal theory and offer only an inadequate response to moral skepticism that permits certain kinds of unethical behavior.

CHAPTER FOUR

Realism and Moral Skepticism

Introduction

In the foregoing chapters I have argued that the constructivist's response to moral skepticism requires us to abandon the idea of deliberate immoral action, despite the fact that we so often interpret ourselves and others as deliberately choosing to do things that are wrong. I have also suggested that there is an in-principle reason why constructivism *cannot* make sense of deliberate immoral action: when a person *chooses* to act a certain way, he takes that course of action to be in some way better than alternative courses of action,[1] but according to constructivism value is the result of the agent's correctly following the procedures constitutive of making a choice. Failure to follow these procedures thus entails a failure to choose.[2] Since any adequate response to moral skepticism will have to show that correct deliberation requires us to act morally, the constructivist ends up maintaining that an agent *cannot* choose to act wrongly. Thus, I claim that making sense of deliberate immoral action requires us to insist that the

[1] Here I am standing in the tradition that claims that we always choose under the aspect of the good. I will not engage the recent literature contesting this point. For discussions for and against this position see Sergio Tenenbaum, ed., *Desire, Practical Reason, and the Good* (New York: Oxford UP, 2010).

[2] I have tailored my objection here to Korsgaard's version of constructivism. Any form of constructivism, however, is going to maintain that value results in some way from the agent's endorsement and will thus be susceptible to a similar objection. After all, to respond to the moral skeptic, the constructivist must maintain that morally permissible actions are most choice-worthy. But choice is a form of endorsement. Indeed, it seems like a particularly strong form of endorsement. Thus, it is doubtful that actions that are not morally permissible could be chosen.

goodness or value of an action does not result from the deliberative processes of the agent. Rather, I maintain that we must be moral realists.

But what can moral realists say to moral skeptics? If the goodness or value of an action does not result from the deliberative processes of the agent, what claim can that value have on the agent? Korsgaard rejects moral realism partly because she believes moral realism is a way of *not* answering moral skepticism by simply insisting that you ought to act in a certain way. If the moral law is a fact existing out there in the world independently of the mind, Korsgaard worries that there is simply nothing the realist *can* say in the face of moral skepticism. What more can the realist do than keep pointing to the 'fact' of the moral law that the moral skeptic questions?[3] If Korsgaard is right in her criticism of realism, it may be that we need to choose between responding to moral skepticism and providing an account of deliberate immoral action.

In this chapter, I will demonstrate that realists do have resources to respond to moral skepticism. Indeed, I will go so far as to maintain that realists are able to offer responses just as compelling as the one Korsgaard offers herself. In the process of making this argument, I will have to consider how best to formulate the normative question and what it means to respond to moral skepticism. Though I do not think Korsgaard clearly demands such a response, I believe some are tempted to think that an adequate response to moral skepticism will compel agents to act morally. I will thus conclude by arguing that the sort of response I outline, though it will not compel agents to act morally, is perfectly adequate and all we can reasonably demand from an ethical theory.

---

[3] See Korsgaard, *The Sources of Normativity*, 30-32.

Clarifying Korsgaard's Challenge

I want to defend the view that values and reasons[4] exist independently from human agents. Two points need to be clarified before I can proceed: what this position requires me to maintain and what, precisely, Korsgaard's challenge to it is. Realism can take many forms, so it is important to carefully identify the position I want to defend. In particular, it is tempting to interpret realism as positing the existence of "queer"[5] metaphysical entities. Korsgaard herself talks at times as if her target claims that moral properties are ontological entities.[6] But this position is something of a straw man. As R. Jay Wallace observes, this interpretation of moral realism "is more often adopted by opponents than by proponents of moral realism."[7] But Korsgaard doesn't *only* reject views of realism that match this description. Korsgaard rejects any ethical view that posits the mind-independence of values and reasons. The central point of her constructivism, remember, is that these are constructed by following certain deliberative procedures.[8]

---

[4] Korsgaard treats reasons and values the same way – as constructions – and understandably so. Korsgaard, like many other philosophers, takes reasons to be normative, so constructing a reason is a way of constructing value. Since I am interested in responding to moral skepticism by providing *reasons* for living morally, my focus in this chapter will be on reasons.

[5] I am, of course, borrowing Mackie's famous language. See J. L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin, 1977), 38.

[6] See for example, Korsgaard, *The Sources of Normativity*, 33, where she describes realism as a commitment to the existence of "intrinsically normative entities."

[7] R. Jay Wallace, "Normativity and the Will" in *Normativity and the Will* (New York: Oxford UP, 2006), 71-81, especially p. 72. For examples of philosophers who accept the mind-independence of normative facts without accepting the existence of metaphysically queer entities, and to see how they defend the claim that moral realism need not entail belief in such entities, see Wallace and T. M. Scanlon, "Being Realistic About Reasons" (Locke Lectures, Oxford University, 2009), http://www.philosophy.ox.ac.uk/lectures/john_locke_lectures/past_lectures (accessed September 13, 2012).

[8] Korsgaard claims that moral realism is "a view about *why* propositions employing moral concepts may have truth values." Namely, it is the view that "propositions employing moral concepts may have truth values because moral concepts describe or refer to normative entities or facts that exist

But the claim that reasons are mind-independent also needs clarification. T. M. Scanlon points out that there are ways in which reasons are *obviously* dependent on human agents. After all, even if we think of reasons as mind-independent facts of some sort or another, they are reasons *for* this agent or that agent. This relational character of reasons makes them, in a certain sense, dependent on human agents since they could not be reasons *for* a given rational agent apart from that agent. Additionally, and relatedly, it is sometimes said that there would be no reasons if there were no rational agents. In response to this claim, Scanlon rightly remarks, "Certainly there would be no point to talking about reasons in the absence of rational agents for whom these are reasons. But it could still be true of certain facts that if there were agents (in the relevant circumstances) then these facts would be reasons for them, if they existed."[9] He concludes, "The kind of independence that normative facts are supposed to have, and do have on my view, comes to this: they are facts that we, individually, could be mistaken about."[10]

But Korsgaard is quite emphatic that her view preserves this feature of moral realism. And Scanlon acknowledges as much, affirming that one of the strengths of Korsgaard's view is that it can "explain what makes claims about reasons correct when they are correct: they are correct if they do indeed follow from requirements of rationality. We can know what reasons we have, on such an account, because we can know what rationality requires."[11] Nevertheless, Scanlon rejects Korsgaard's view of reasons. How then are we to distinguish the mind-independence of reasons that Scanlon

---

independently of those concepts themselves." Korsgaard, "Realism and Constructivism in Twentieth-Century Moral Philosophy," 302.

[9] Scanlon, "Being Realistic About Reasons," Lecture 5, 14.

[10] Ibid.

[11] Scanlon, "Being Realistic About Reasons," Lecture 1, 13.

and I want to preserve from Korsgaard's position? Recall the discussion in my first chapter in which I observed that Korsgaard can maintain that agents are mistaken about whether they *have* reasons, but she has to deny that agents might act on reasons they mistakenly take to outweigh or override other considerations. I believe this insight may help us identify the way in which we can be mistaken about our reasons that is distinctive of views that take reasons to be mind-independent.

While Korsgaard can allow that we may think we have reasons when we actually do not, she cannot allow for other sorts of mistakes that occur, for example, when: 1) we really do have reasons for our action, but those reasons ought to be overridden by some other consideration(s) we fail to give proper weight, or 2) when there are reasons we are unaware of because we lack the relevant desires. These two sorts of mistakes involve making judgments about what reasons there are prior to our deliberations, while Korsgaard is restricted to reflecting on the products of our deliberations. When we accuse agents of failing to weigh reasons properly, we are talking about reasons that ought to figure in deliberations – not reasons that result from these deliberations. Similarly, when we countenance the idea that an agent may be unaware of his reasons, we suggest that those reasons exist whether they enter into the agent's deliberations or not. Thus, it seems to me that the correct way to characterize the mind-independence of reasons is to say both that we can be mistaken about them and that they precede deliberation. This is all that I claim is needed to make sense of deliberate immoral action, and it is all I intend to defend.

What precisely, then, is Korsgaard's concern about this view's ability to respond to moral skepticism? Korsgaard seems to be concerned that such normative facts will not

be able to motivate skeptics to act accordingly. But Scanlon suggests two different ways of interpreting this concern. On the one hand, Scanlon observes, this worry could be construed as a version of Hume's objection that the belief that the world is a certain way cannot motivate an agent to act one way or another unless the agent has some relevant desire. That is, apart from some relevant desire, a belief about the world *cannot* provide the agent with a reason to act. Alternatively, Scanlon believes Korsgaard may be asking how it is that we could show an agent that he should *accept* that a given fact is a reason to act a certain way. Even if we grant that facts *are* reasons, if the agent denies that this particular fact is a reason to act a certain way, what could we say to him? How can we get an agent to see that *this* fact is a reason and thus be motivated to act accordingly? Scanlon believes Korsgaard's normative question is probably raising the latter concern. As noted in my first chapter, Korsgaard certainly shares the former objection. But if her normative question is simply a version of the Humean objection, why does she not treat it as such?

I believe we can get a clearer sense of what Korsgaard's concern is by looking at the sort of answer that she proposes. In *The Sources of Normativity*, her argument seems, roughly, to run as follows: insofar as I care about anything, I am rationally committed to morality. And her argument in *Self-Constitution* suggests that insofar as I am an agent I am rationally committed to morality. The point to notice is that, in both cases, Korsgaard suggests that all normally-functioning human beings (since most of us are agents who care about *something*) are committed to morality on pain of irrationality. Failure to act morally betrays some kind of internal inconsistency. Korsgaard's proposed responses to the normative question suggest that she wants to know whether morality's authority over

us is a matter of purely formal rational constraint. That is, can it be shown that it would be inconsistent of us to act immorally?

Perhaps Scanlon is right to suggest that Korsgaard's normative question, when posed to a realist, is best interpreted as the question of whether we can say anything to the person who does not recognize that a certain fact is a reason. But we can make Korsgaard's normative question clearer still: can a realist show that an agent must act morally on pain of irrationally? While Korsgaard believes her constructivist approach can demonstrate that immorality involves agents in rational inconsistency, she doubts that realism can do so. This is where her worry about the motivational power of facts (or lack thereof) becomes relevant. If facts cannot motivate, it is hard to see how an agent could be involved in a formal rational inconsistency simply because that agent failed to respond to certain facts that have no inevitable motivational pull on the agent. Thus, I will begin by considering the Humean objection that facts cannot motivate. I will then consider whether realists can respond to Korsgaard's normative question as I have just interpreted it and argue that they can.

Responding to the Humean Objection

I will begin this section by first asking why we should think facts *cannot* motivate. It is worth noting that I will alternate between describing reasons as facts and describing them as beliefs. Doing so should have no effect on my argument. In keeping with the distinction I made between objective and subjective reasons in my first chapter, reasons are facts insofar as reasons are objective. Subjective reasons, however, are beliefs. If objective reasons are facts, it is possible (in principle) for agents to form beliefs concerning those facts. The question is whether those beliefs can motivate agents

to act. I will then turn my attention to critiques of the understanding of reasons typically offered as an alternative to accounts that identify reasons with facts or beliefs. Most people who object that beliefs cannot motivate action propose that reasons for action must be grounded in some desire(s) of the agent. Korsgaard's own view is more nuanced than a simple desire-based account of reasons, so I will also discuss how her view differs from these more naïve desire-based theories. Her account, however, is still agent-based. As a result, some of the objections to desire-based theories will likewise apply to her account of reasons. I will thus survey reasons to reject both desire-based views and Korsgaard's view of reasons.

Why should we think that beliefs cannot motivate? If we consult our experience, it seems that we are often motivated by the belief that something would be good for us to do, or that something would be bad for us to do. At least *prima facie*, then, it is not clear why we should assume that facts can*not* motivate. Scanlon proposes that the reason beliefs can motivate human action is that what it *is* to be rational is to respond to reasons. He writes,

> A rational agent is, first, one that is capable of asking questions about the reasons he or she has for performing certain actions or for holding various other attitudes. Second, a being is a rational agent only if the judgments that it makes about reasons make a difference to the actions and attitudes it proceeds to have.[12]

While this explanation should not be taken as further evidence that beliefs can motivate action, it seems like a plausible interpretation of our experience: we recognize certain facts as giving us reasons to act a certain way and are thus moved to act accordingly. In light of our experience, what reason do we have to insist that reasons *must* depend on some prior conative state of the agent's?

[12] Scanlon, "Being Realistic About Reasons," Lecture 3, 3.

The closest Korsgaard comes to giving us a reason for thinking that facts cannot motivate follows a comparison she draws between knowledge of facts and the possession of a map. Knowledge, and therefore belief, in this metaphor requires nothing more than possession of the map. Action, however, requires the agent to *use* the map. But possessing the map can never by itself prompt the agent to use the map. By extension, simply knowing what makes an action good will never motivate a person to act accordingly unless that person is already committed to performing good actions – that is, to using or following norms that govern good action. Korsgaard writes, "If it is just a fact that a certain action would be good, a fact that you might or might not apply in deliberation, then it seems to be an open question whether you *should* apply it."[13] The most obvious way a realist might explain the motivational power of facts, she thinks, is through what she calls "the instrumental principle."[14]

The instrumental principle instructs the agent to take the means to his ends. If my end is to have an enjoyable evening, the fact that I enjoy movies may motivate me to take in a film. The problem, Korsgaard maintains, is that the realist cannot coherently explain the authority of the instrumental principle itself. Realists will want to maintain that "you ought to take the means to your end" is a fact that motivates action, or a reason. But realists cannot explain how that fact motivates action without assuming that the agent is already motivated by the instrumental principle.

This argument strikes me as unfair to realists. First of all, the metaphor of the map seems to overlook that realists believe reasons belong to a distinctive class of facts

[13] Korsgaard, "Realism and Constructivism in Twentieth-Century Moral Philosophy," 317.

[14] "The obvious way to understand how facts motivate us is by means of a kind of extension of the instrumental principle itself." Korsgaard, "Realism and Constructivism in Twentieth-Century Moral Philosophy," 317.

that necessarily motivate rational agents when recognized as reasons. Not all facts are reasons. By comparing the knowledge of reasons to the possession of a map you might not use, Korsgaard fails to take seriously the realist view of reasons. Secondly, why should we expect *anyone* to be able to explain why we ought to take the means to our ends? Consider Scanlon again. According to Scanlon, a reason is "a consideration that counts in favor of" something.[15] That something is a means to fulfilling my end seems an especially clear example of how a fact might count in favor of my acting a certain way. And if Scanlon is right in claiming that what it is to be rational is to respond to reasons, then the person who does not respond to the fact that something is a means to his end is irrational. There is not much use in reasoning with an irrational person, whether your account of reasons is realist or constructivist.

It might be objected that Korsgaard is not concerned so much to *persuade* irrational people to adhere to the instrumental principle as she is to *explain* the fact that rational people *do* adhere to it. Her argument still strikes me as unfair insofar as she accuses the realists of incoherence. I do not see anything incoherent in the Scanlon-inspired account I sketched in the paragraph above. What Korsgaard might object to is that Scanlon leaves unexplained something she feels demands explanation: that rationality requires us to respond to certain considerations. This objection is hard to evaluate. Does the fact that rational agents adhere to the instrumental principle need some additional explanation besides the fact that they are rational? How do we determine whether an explanation is required? While it is not immediately obvious to me that an

[15] T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Belknap Press, 2000), 17.

explanation is needed, if Korsgaard *is* able to offer a compelling explanation, it may count as a strength of her constructivist position.

But I do not believe Korsgaard provides any more of an explanation than the one I sketched above. Korsgaard believes that the instrumental principle is, along with the categorical imperative, one of the constitutive principles of agency.[16] In other words, insofar as we are rational agents, we just do adhere to the instrumental principle. And remember, Korsgaard observes that we cannot escape our agency. Our adherence to the instrumental principle thus seems to follow simply from our status as rational agents. This conclusion can also be seen by reflecting on Korsgaard's understanding of the categorical imperative's universalizability requirement. As observed in my previous chapter, Korsgaard has promoted an interpretation of the universalizability requirement that precludes agents from acting on maxims that, when universalized, involve the agent in a practical contradiction. A practical contradiction occurs when the universalization of a maxim would prevent the agent from realizing her end.[17] In other words, a practical contradiction runs afoul of the instrumental principle. Since the categorical imperative is also a constitutive principle of agency, it again seems that the explanation for agents' adherence to the instrumental principle is simply that rational agency requires it.

Since I see no reason to demand an explanation for agents' adherence to the instrumental principle besides the fact that rationality requires that adherence, and since Korsgaard seems unable to provide a more nuanced explanation, I do not find her argument that realist accounts of the authority of the instrumental principle are incoherent

[16] In fact, she maintains that the efficacy captured by the instrumental principle and the autonomy of the categorical imperative are two sides of the same coin: agency. See Korsgaard, *Self-Constitution*, Chapter 5.

[17] See Korsgaard, "Kant's Formula of Universal Law."

compelling. Korsgaard provides us no further reason for thinking that facts cannot motivate. In fairness, there may be reasons to think that facts cannot motivate that Korsgaard does not articulate. Derek Parfit has suggested that the primary motivation for thinking that reasons must rely on the conative state of the agent is the result of certain metaphysical commitments. Namely, those committed to naturalist metaphysics may worry that the normativity of reasons cannot be explained in natural terms unless that normativity has something to do with the conative state of the agent.[18] Normativity, according to both Parfit and Scanlon, is an irreducible feature of those facts that serve as reasons.

It is not clear that Korsgaard is opposed specifically to non-naturalist metaphysics, but she is clearly hostile to the idea that metaphysics might play any helpful role in thinking about ethics. This hostility seems to have its source in two worries: first, realist metaphysics could involve ontologically 'queer' entities and, second, any factual claim about the world could not possibly be the right sort of thing to be a reason. I have already observed that the worry that realism commits one to belief in the existence of ontologically queer entities is something of a straw man, though I do not intend to undertake a thorough discussion of realist responses to this sort of metaphysical worry.[19] Korsgaard's second worry seems to me simply to assume that facts cannot motivate, contrary to what our experience, *prima facie*, suggests. I will thus proceed on the

[18] Derek Parfit, *On What Matters*, ed. Samuel Scheffler, Vol. I (New York: Oxford UP, 2011), 110. In that same passage, Parfit offers a response to those motivated by such naturalist worries.

[19] For Scanlon's discussion of the matter see Scanlon, "Being Realistic About Reasons," Lecture 3, 16-17. See also Parfit, *On What Matters*, Vol. I, 110.

assumption on that we have no reason to think that facts cannot motivate and begin my critique of desire-based theories of reasons.

According to desire-based theories of reasons, all reasons for action have some basis in the inclinations and desires of the agent. That is, there can be no reason for action where there is no corresponding desire on the part of the agent. Nagel was among the first contemporary ethicists to challenge this view, remarking that desire-based accounts typically fail to distinguish between what he calls 'motivated' and 'unmotivated' desires. A motivated desire is a desire *for which the agent has a reason*. An unmotivated desire, by contrast, is not had for any particular reason. The basic idea is that some desires are arrived at by deliberating about what we should do or what we should want. In this way, some desires are acquired and some desires are overruled. It is trivially true, Nagel thinks, that all actions must involve either motivated or unmotivated desires. But if an action involves a *motivated* desire, that desire cannot provide the reason for the action, since the desire to perform that action was itself acquired for a reason – standardly the reason for the action. Nagel explains,

> The claim that a desire underwrites every act is true only if desires are taken to include motivated as well as unmotivated desires, and it is true only in the sense that *whatever* may be the motivation for someone's intentional pursuit of a goal, it becomes in virtue of his pursuit *ipso facto* appropriate to ascribe to him a desire for that goal.[20]

The claim that reasons involve desires is thus neither substantive nor interesting.

But what if someone were to take the stronger line that all reasons for action depend on some *unmotivated* desire the agent has? While it would be hard to deny that there are motivated desires which we acquire for a reason, perhaps that reason is itself provided by an unmotivated desire. This position has some implausible implications.

[20] Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton UP, 1970), 29, italics his.

Nagel illustrates one such problem with an example of a person putting a dime (yes, just a dime) into a vending machine slot to get a soda to quench his thirst. While it might seem plausible that his thirst explains why he drinks a soda, it is less clear how it motivates him to put a dime into a slot. Even if, Nagel claims, we supplement the desire to drink with the relevant knowledge, like the knowledge of how vending machines work, it is still mysterious how the desire to *drink* combines with this knowledge to move the agent to put a dime in the slot. Nagel cautions his readers against attempts to supplement the desire to drink with a further desire to put a dime in the slot because this further desire requires explanation. Otherwise, what would prevent a person from acquiring the desire to put a dime into his pencil sharpener in response to a desire to drink?

Think of the desire to put a dime in the slot of the vending machine as a motivated desire acquired after deliberation on the fact that one is thirsty and that one way to satisfy this thirst is to put a dime in the slot of the vending machine. If we try to explain the acquisition of this motivated desire in terms of some unmotivated desire, the connection between the motivated desire to put a dime in the slot and the unmotivated desire to drink needs to be explained. Why does the person in this example acquire the *specific* desire to put a dime into the vending machine rather than, say, the desire to put a dime into a pencil sharpener? If we appeal only to unmotivated desires, there *is* no reason why you should acquire the desire to put the dime into the vending machine rather than the desire to put the dime into a pencil sharpener. It thus seems that it makes just as much sense to put a dime into the pencil sharpener as it does to put the dime into the vending machine.

And it seems implausible that putting a dime into a pencil sharpener could be just as intelligible as putting a dime into a vending machine because one is thirsty.[21]

The solution to the problem of explaining how the desire to drink can translate into a reason to put a dime in the slot is to posit a general principle according to which having a reason to pursue an end entails having a reason to take the means necessary to attain that end. Having reason to drink thus entails having a reason to put a dime in the slot because putting a dime in the slot is the means to attaining the end of taking a drink. The incorporation of this principle into a desire-based account of reasons will look something like this: an agent has a reason to take the means to an end for which the agent has a present desire. But this account, Nagel argues, does not adequately explain prudential actions which aim at the satisfaction of future desires.

The desire-based account maintains that agents who act prudentially must have a present desire for the satisfaction of future desires. Otherwise, the desire-based account could not explain why those agents have reason to act in ways that best serve the satisfaction of those future desires. Present desires must *always* serve as the source of reasons. But this view again has implausible consequences. For example, Nagel observes, I may presently desire something for the future that I know I will not desire in the future. But in that case, it follows that I have reason to prepare to do something even though I know I will have no reason to do it when the time comes.[22] And such cases are not hard to imagine. I may presently want to go to the gym this afternoon, and so I now have reason to clear my schedule so that I will be able to do so. But I may also know that

[21] See Nagel, *The Possibility of Altruism*, 33-34.

[22] See Nagel, 39. I will not survey all of Nagel's arguments against this desire-based account of reasons. For other arguments, see Nagel, 39-46.

I become rather indolent and lazy in the afternoons, such that I will cease to want to go to the gym when this afternoon rolls around. On the desire-based view of reasons, I will cease to have a reason to go to the gym this afternoon, but I nonetheless have a reason *now* to clear my schedule so that I can go to the gym this afternoon.

Attempts to explain motivated desires in terms of unmotivated desires thus fail. More recently, Derek Parfit has also pressed desire-based theories on problems that arise when they try to explain what it means to have a reason for a desire. He presents what he calls "The Agony Argument": We all have reason to avoid future agony, but desire-based theories of reasons imply that we have no such reason. Therefore, desire-based theories are false.[23] The idea is that we can imagine someone who simply lacks the desire to avoid future agony and thus, according to desire-based theories, does not have a reason to avoid that future agony. Someone might object, however, that agony would not be agony unless the agent did *not* want to be in that state. Thus, we *cannot* imagine someone who lacks the desire to avoid future agony. Parfit amends his argument, maintaining instead that desire-based theories cannot maintain that all agents have a reason to *want* to avoid future agony.[24] While it might plausibly be claimed that agents must, by definition, desire to avoid future agony, it does not follow that agents must desire to have that desire. But if we accept a desire-based theory, then the agent's lack of a present desire to want to avoid future agony means that the agent has no reason to want to avoid future agony. Parfit's new argument runs thus: We all have reason to want to avoid future agony, but

[23] Parfit, *On What Matters*, Vol. I, 78.

[24] Ibid., 74.

desire-based theories imply that we have no such reason. Therefore, desire-based theories are false.[25]

Both Parfit's and Nagel's challenges, I believe, capitalize on the normative nature of reasons. That is, if something is to count as a reason, it must direct you to, for instance, act a certain way or form a certain belief. A reason tells you what you *ought* to do, believe, feel, etc. The problem with the person who puts a dime into a pencil sharpener because he is thirsty is that he is misguided, as is the person who has no desire to avoid future agony. Korsgaard, however, is sensitive to the normativity of reasons. Indeed, if you recall our discussion in chapter one, she rejects noncognitivist views precisely because they deprive reasons of their normative force. Thus Korsgaard believes reasons *involve* desires, but are not simply grounded in desires.

Korsgaard's view is what Scanlon refers to as a 'rationality-based' account of reasons.[26] Desires only provide the agent with reasons to act when the action proposed by the desire passes the test of rationality, that is, the test of universalizability. But although Korsgaard emphasizes rationality rather than desires, she still does not adequately account for the normativity of reasons. Much of this dissertation has highlighted Korsgaard's difficulty by pointing to her failure to demonstrate that the life of the Idealized Mafioso is not choiceworthy. But there is another way in which Korsgaard's account fails to provide the agent with adequate guidance. Namely, it is not clear how Korsgaard could explain how an agent might be directed to acquire a desire she does not currently have.

---

[25] Ibid.

[26] Scanlon, "Being Realistic About Reasons," Lecture 1, 11-16.

According to Korsgaard, desires give us candidate reasons. Those candidates that pass the test of rationality *are* reasons. But what if there is no desire to, say, avoid future agony? In that case, there is no candidate reason to run through the test of rationality. Without that desire in place, there can be no reason to avoid future agony – let alone a reason to want to avoid future agony. How does someone acquire reasons for desires she does not already have on Korsgaard's account? We might have reasons to acquire desires as means to our ends. If, for example, I desire to get in shape, that desire might make me wish that I wanted to run. But the desire to get in shape has to actually give rise to a desire to want to run. Once I have the desire to want to run, I may have reasons to take actions that cultivate the desire to run, but unless I have the desire to want to run, I have no reason to want to run or to take actions to cultivate that desire. And this seems wrong. Surely it would be a good thing for me to want to run as a means to getting in shape, whether I have the desire to want to run or not.

Furthermore, Korsgaard does not give us any way of evaluating the desires we have. Running maxims through a universalizability test, at most (if Korsgaard is right), tells us whether we can act on those maxims, but it does not tell us whether we should desire the end identified in the maxim or cultivate a desire for something else entirely. Talk of wanting to get in shape and cultivating the desire to run is fairly innocuous, but let us suppose that I have the desire to kill my brother. It may be that I cannot universalize any actions proposed by that desire, but the test of universalizability tells me nothing about the desire itself. This inability to offer an account of how to evaluate desires is a serious flaw in a theory of normativity. We need guidance not only in how we act, but in how we feel as well. One of the appeals of Kant's ethics is that he resists

grounding our knowledge of moral obligations in feelings, among other reasons, *because* we do not feel as we ought. Nevertheless, Kant enjoins us to pursue appropriate desires.[27] But on what grounds do we evaluate desires?

This concern about how we acquire reasons for desires or evaluate desires is closely related to Scanlon's objection to rationality-based theories. He observes that desire- and rationality-based accounts of reasons have most appeal when we consider cases in which one person tries to persuade another to act a certain way. When trying to convince another person to act a certain way, it is often most effective to appeal to that person's present desires, or to rational consistency in the pursuit of those desires. But when one deliberates with oneself, it is not so clear that those sorts of considerations tell the whole story. Scanlon observes that the problem is that "[o]ne can always ask oneself whether why [sic] one should have these attitudes in turn."[28]

The ironic thing, he notes, is that Korsgaard is keenly aware of the agent's need to deliberate with herself, even stressing the importance of asking oneself why one should have certain attitudes and desires until it is "impossible, unnecessary or incoherent to ask why again."[29] But he also points out that Korsgaard's claim "would be much less plausible without the disjunct 'unnecessary.'"[30] Few people would automatically accept the claim that we should ask why we should have certain desires until it is either impossible or incoherent to ask why again. It seems to make more sense that we should

---

[27] See, for example, Kant's discussion of moral enthusiasm in Kant, *KpV*, 5:85-5:86.

[28] Scanlon, "Being Realistic About Reasons, Lecture 1, 20.

[29] As quoted in Scanlon, "Being Realistic About Reasons," Lecture 1, 20. Cf. Korsgaard, *The Sources of Normativity*, 33.

[30] Scanlon, "Being Realistic About Reasons," Lecture 1, 20.

keep asking why until it is unnecessary to ask again. But what makes it unnecessary to ask again, if not substantive claims about the reasons one has?[31]

We might be tempted to think that Scanlon isn't fair to Korsgaard when he assumes that only substantive claims about reasons can make it unnecessary to ask why one should have certain attitudes and desires. Korsgaard might, for example, invoke reflective equilibrium: you no longer need to ask why you should have certain attitudes when, after reflection, you have successfully resolved any tension between your various beliefs and attitudes. Given Korsgaard's emphasis on the construction of a unified self, such a maneuver seems fitting. But this maneuver also seems unsatisfying. Reflective equilibrium can be sought simply by appealing to an agent's present desires and rational consistency, but such an appeal hardly seems to provide sufficient guidance. After all, when an agent is trying to decide for herself what attitudes and desires she should have – like whether she should care about future agony – appeals to her present desires and rational consistency will get her nowhere if she currently lacks the desire to avoid future agony. To borrow Korsgaard's terminology again, since there is no candidate reason, there cannot be a reason on desire- and rationality-based accounts. An agent who is trying to decide whether she should care about future agony will need to look elsewhere for considerations to guide her deliberations.

Responding to Korsgaard's Normative Question

Even if we grant, though, that facts can motivate and that there is no reason to accept an alternative, problematic account of reasons, Korsgaard's normative question may still present a challenge to proponents of realist metaethics if realists cannot show

[31] Ibid., 20-21.

that agents are rationally required to live morally. In this section, I will first show how Korsgaard's proposed response to moral skepticism does not actually improve upon responses offered by moral realists. Indeed, realists are able to meet Korsgaard's challenge as effectively as she does. But some may worry that these responses are not effective *enough* insofar as it is not clear that their responses will compel agents to act morally. I will thus proceed to argue that we do not *need* to offer a stronger response. There are good reasons to give up on finding a response to moral skepticism that will persuade anyone to act morally.

I want to begin by observing that there is no in-principle reason why a realist could *not* construct a response to moral skepticism that appealed to the requirements of reason. After all, if what it means to be rational just *is* that one responds appropriately to reasons, a realist needs merely to show that reasons to act morally always outweigh or override reasons to act otherwise in order to show that an agent would be rationally constrained to comply. There are probably several ways a realist could do this. One might, for instance, attempt to show that morality preserves some essential feature of human nature or rational engagement that thus always gives moral reasons a trump card over any other considerations. I will discuss one such realist maneuver momentarily.

But even if an agent has most reason to act morally, a realist must grant that the agent can deny that fact or fail to acknowledge the significance of that fact, much as an agent can fail to grasp how a certain fact conflicts with a belief he has formed. The distinction I made between objective and subjective reasons may again be helpful here. Insofar as we have been trying to account for the normative force of reasons, I have largely been referring to objective reasons. But failure to recognize these reasons as

reasons means that the agent may have no subjective reason to act accordingly. It is plausible that Korsgaard's appeal to rational necessity is meant to help us avoid this problem. Simply in virtue of reasoning consistently, we should be able to show agents that the only reasons they have are reasons to act as morality requires. I am not convinced, however, that we should construe Korsgaard's appeal to rational necessity as designed to accomplish this task of persuading agents to act morally. Though I am primarily concerned to show how realists might respond to Korsgaard's normative question, interpreted as a demand that morality be shown to be rationally required of agents, some may be inclined to expect a response to moral skepticism that *persuades* others to live morally. Thus, a secondary goal of this section will be to address this sort of demand.

I want to begin by drawing a comparison between Korsgaard's response to moral skepticism and a response suggested by Nagel. In *The Sources of Normativity*, Korsgaard cites Nagel as an example of a contemporary realist who refuses to answer the moral skeptic, claiming that his response to the moral skeptic is "an expression of confidence and nothing more."[32] But Korsgaard refers to Nagel's work in *The View from Nowhere*, in which he defends the possibility that there are mind-independent normative facts. He is *not* illustrating how we might demonstrate that agents are rationally required to act morally. To someone who doubts that there are mind-independent normative facts he writes:

> I think the burden of proof has been often misplaced in this debate and that a defeasible presumption that values need not be illusory is entirely reasonable until it is shown not to be… [A] lot depends on whether the possibility of realism is

[32] Korsgaard, *The Sources of Normativity*, 41.

> admitted in the first place. It is very difficult to argue for such a possibility, except by refuting arguments against it.[33]

Towards the end of her first lecture in *Sources* Korsgaard seems to acknowledge that these are different issues, but then quickly slides from one to the other:

> The realist's belief in the existence of normative entities is not based on any discovery. It is based on his *confidence* that his beliefs and desires are normative. So even if it is true, the realist cannot answer the normative question. But why should this matter? If confidence can support a metaphysics which in turn is supposed to support the claims of morality, why can't confidence support the claims of morality more directly?[34]

Her suggestion is that the moral realist assumes that if confidence is an adequate ground for the adoption of a certain metaphysical position, it may serve as an adequate basis for a response to the normative question. But the idea that the metaphysics of moral realism is supported merely by confidence strikes me as uncharitable, especially in light of the considerations we just adduced for thinking that reasons may be irreducibly normative facts. In addition, I worry that Korsgaard is too quick in extending her criticism of the realist's view of the nature of normativity to the realist's response to a person who does not recognize a reason – or, perhaps, overriding reason – to act morally.

Nagel offers a nuanced response to the person who does not recognize a reason to act morally in *The Possibility of Altruism*. At first glance, it might not look like Nagel provides much of a response. He tells us openly that he believes "no form of scepticism, whether epistemological or moral, can be shown to be impossible."[35] But he immediately continues, "The best one can do is to raise its cost, by showing how deep and pervasive

---

[33] Thomas Nagel, *The View from Nowhere* (New York: Oxford UP, 1989), 143. Korsgaard herself cites part of this passage in her description of Nagel's position. See Korsgaard, *The Sources of Normativity*, 41, n.67.

[34] Korsgaard, *The Sources of Normativity*, 48.

[35] Nagel, *The Possibility of Altruism*, 143.

are the disturbances of thought which it involves."[36] Clearly, then, Nagel intends to say *something*, and it is more than just a matter of insisting that you should do *X* because you should do *X*.

In fact, Nagel suggests two ways a realist might respond to the moral skeptic. First, he observes that he might simply show that moral skepticism "conflicts radically with practical intuitions to which everyone, the sceptic included, is chronically subject."[37] But Nagel himself opts instead to argue that moral skepticism requires the abandonment of even deeper commitments than practical intuitions. After all, as he points out, those intuitions themselves may be challenged by skepticism. Nagel argues, rather, that motives to act morally[38]

> depend on the fact that our reasons for action are subject to the formal condition of objectivity, which depends in turn on our ability to view ourselves from both the personal and impersonal standpoints, and to engage in reasoning to practical conclusions from both of those standpoints. These are forms of thought and action which it may not be in our power to renounce.[39]

He believes that to be a moral skeptic requires you to abandon the ability to see yourself both from a personal point of view and from an impersonal point of view which takes you to be one person among others.

Nagel assumes that "we are not solipsists."[40] What he means by this is that we do not take our experiences to be fundamentally different from others'. When someone else

---

[36] Ibid.

[37] Ibid., 144.

[38] What Nagel actually refers to are "altruism and related motives," but as he mentions them in connection with responding to moral skepticism, I take it that it is fair to describe these as "motives to act morally."

[39] Nagel, *The Possibility of Altruism*, 144.

[40] Ibid., 106.

complains of a headache, we take it that this person is experiencing something similar to the dull, throbbing pain we experience when we have headaches. Given that our experiences are *not* fundamentally different from others' experiences, anything that counts as a reason *for me* will also count as a reason for *anyone* in those circumstances. Thus, he suggests that we can view the entire world impersonally – in terms of the circumstances each person finds himself in and the reasons that person has to act in certain ways. The personal point of view locates my position in the world. For example, I can describe my current situation and reasons impersonally like so: Heidi Giannini has reason to spend the afternoon writing. But I simultaneously recognize that *I* am Heidi Giannini and, thus, that *I* have reason to spend the afternoon writing. If I fail to take up *both* the personal and impersonal standpoints, I will either fail to explain how judgments can motivate (since the recognition that Heidi Giannini has reason to act a certain way does not necessarily provide me with any corresponding reason to act) or I will fall into solipsism by failing to recognize that the reasons *I* have are reasons *anyone* would have in those same circumstances.[41]

The details of his argument are not important for my purposes. For now, what *is* important for my purposes is the nature of his response: the fact that his argument against moral skepticism claims that moral skepticism requires us to abandon some facet of human existence that is, at the very least, difficult to escape. Of course, he concedes that the skeptic might ask why he *should* occupy both the personal and impersonal points of view. And he's not sure there's any way to argue with such a person: "The same data which when viewed personally justify a practical judgment will not do so when viewed

[41] See chapters 11 and 12 of Nagel, *The Possibility of Altruism*.

impersonally. What further objection can be raised if someone remains unabashed about coming apart in this way?"[42] But he also seems to doubt whether it is psychologically *possible* for people unabashedly to come apart in this way. Hence he claims that "it may not be in our power to renounce"[43] the tendency to occupy both points of view. We just *aren't* solipsists.

In the last chapter I pointed out that Korsgaard's response to the moral skeptic depends on the publicity of reasons which, in turn, depends on the fact that human beings simply *do* value one another – at least enough to treat the reasons of others as providing themselves with reasons to act. And Korsgaard seems to think that it is acceptable to appeal to this fact about human beings because she takes it that treating others' reasons in this way is practically inescapable for normal human beings, as a matter of psychological fact. That is, for the most part you cannot *help* but treat others in this way. Notice, then, that Nagel's response to the moral skeptic appears to follow the same form as Korsgaard's: in order to be a moral skeptic, you must give up something it is doubtful you *can* give up. More to the point, while it is true that the realist could say nothing to the skeptic who doubts whether he must occupy the personal and impersonal points of view, Korsgaard is in a similar position with respect to the skeptic who doubts whether he must value humanity. Korsgaard accuses realists of digging in their heels and refusing to respond to moral skepticism, but as it turns out, they need only do so to the same extent that Korsgaard does.

---

[42] Nagel, *The Possibility of Altruism*, 145.

[43] Ibid., 144.

Of course, as I observed in my last chapter, Korsgaard appears to acknowledge the limits of her response, insofar as she admits that she has no answer for the complete normative skeptic. But she does think she has a response to the moral skeptic. That is, she cannot convince the person who sees no reason to value *anything* to live morally, but she does think she can show the person who values *something* that he is rationally committed to being moral. The argument seems to run something like this: In valuing something, you commit yourself to being a certain sort of person, and thus betray a commitment to value your own humanity. Indeed, when you look at the way you relate to other people and their reasons, it is apparent that you treat their reasons as having some hold on you, further evidence of your commitment to the value of humanity. But in valuing humanity, you are committed to morality. On what grounds can you suppose that you have reason *not* to act morally? But again, Nagel mirrors this response to the moral – as opposed to the complete normative – skeptic: When you reason about how to act, you just *do* take on both the personal and impersonal points of view. But in taking on these two standpoints, you are committed to morality. On what grounds can you suppose that you have reason *not* to act morally?

In order to support her claim that realist responses to moral skepticism are objectionably different from her own, Korsgaard might argue that the sort of maneuver Nagel made in responding to moral skepticism is incompatible with his realism. In "The Reasons We Can Share" she does suggest that Nagel's position "is not fully consistent,"[44] and even argues that at times Nagel seems to approximate her own position rather than a realist's. Once again she draws on his work in *The View from Nowhere*, pointing to two

[44] Korsgaard, "The Reasons We Can Share," 281.

passages she takes to be in tension. In one passage he writes, "The reasons are real, they are not just appearances. To be sure, they will be attributed only to a being that has, in addition to desires, a general capacity to develop an objective view of what it should do. Thus, if cockroaches cannot think about what they should do, there is nothing they should do."[45] In the next he claims that "the pain can be detached in thought from the fact that it is mine without losing any of its dreadfulness. It has, so to speak, a life of its own. That is why it is natural to ascribe to it a value of its own."[46] Korsgaard takes the first passage to be claiming that there are only reasons when there are reflective beings that can have them and the second passage to be claiming that there are reasons existing independently from any agent, or agent-neutral reasons.[47]

On the one hand, Korsgaard sees Nagel claiming that pain is a bad thing all on its own. On the other hand, she sees him claiming that cockroaches cannot have reasons for action if they cannot reflect on their actions. But surely, she points out, a realist must believe that "the pains of animals who cannot think objectively about what they should do must be bad in the same way as the pain of animals who can."[48] Nagel, however, never claims that the pain of cockroaches is not bad – just that there is nothing cockroaches should do, since they cannot reflect on what they should do. The objective badness of pain only motivates action in those who can *recognize* that badness as a reason to act in a certain way. Thus, we attribute reasons to those who reflect on what

[45] As quoted in Korsgaard, "The Reasons We Can Share," 280. Cf. Nagel, *The View from Nowhere*, 150.

[46] As quoted in Korsgaard, "The Reasons We Can Share," 280. Cf. Nagel, *The View from Nowhere*, 160.

[47] Korsgaard does describe herself as defending an intersubjectivist account of agent-neutral reasons. Here I take it to be understood that I am referring to a realist account of agent-neutral reasons.

[48] Korsgaard, "The Reasons We Can Share," 281.

they should do and so are engaged in the process of discovering and recognizing reasons. I see no problem with this position or with Nagel's response to moral skepticism given his realism. I know of no other argument Korsgaard has made that suggests that Nagel's position is contradictory.

Furthermore, it is not clear that this alleged tension in Nagel's thought even matters. Even if Korsgaard is right that Nagel vacillates between realist and constructivist treatments of reasons, she would not have shown that it is impossible for a realist to make the sort of maneuver Nagel proposes in response to the moral skeptic. Why can't a realist endorse the idea that our reasons for action – the reasons that we recognize and act on – are subject to a condition of objectivity? Indeed, this condition of objectivity is strikingly similar to Korsgaard's claim that reasons are public, a feature which she acknowledges may be explained by realist accounts of reasons. Recall that in *The Sources of Normativity*, she observes that there are "two ways to go about showing that reasons are inherently public. One is to try to defend some form of substantive moral realism."[49] I thus conclude that we have no reason to suppose that Nagel's response to moral skepticism is incompatible with moral realism.

The responses offered by Korsgaard and Nagel may thus succeed in showing that agents are rationally required to live morally *given* certain practically inescapable features of human psychology, but it is not clear that the accounts provided by either philosopher will *compel* agents to act morally. And it might be tempting to think we *should* have something more to say to those who do not recognize reasons for moral action or their relative strength. Korsgaard may not have provided us with such a

[49] Korsgaard, *The Sources of Normativity*, 135.

response, but we may still feel that such a response is worth pursuing. I don't think this is right. Scanlon freely admits that, when faced with a person who does not recognize a reason, there comes a point where he must simply insist that it *is* a reason. But he also observes that disagreements about beliefs often end this way too: "To insist, at the conclusion of an unsuccessful attempt to persuade someone, that your normative judgment is correct, [sic] is indeed unhelpful foot stomping. But this is equally true when the disagreement is about some matter of empirical fact."[50] Scanlon does not attempt to provide any sort of response to Korsgaard's normative question, not even a response like the one Nagel formulates. Nevertheless, his insight here seems accurate. Even if we can show, through some philosophical theory, that we are rationally committed to acting morally, it does not follow that we will be able to either convey the rationality of acting morally in some given instance or convince the person we're talking with to *be rational*.

The inability to persuade any particular rational agent that she should act morally *could* be problematic, though, if our inability to do so corresponds to a larger inability to reason together about normative demands *at all*. That is, we might not be able to persuade others to act morally because it might be impossible for us to engage in any meaningful deliberation about what reasons we have. If this were true, it would present a compelling challenge to realists, since we do sometimes seem to engage in cooperative deliberation about moral issues. But I don't think the moral realist's inability to provide this sort of persuasive argument for acting morally is the result of some more fundamental difficulty in explaining how we can deliberate together about our reasons.

[50] Scanlon, "Being Realistic About Reasons," Lecture 3, 21. When he made this comment, Scanlon was not responding directly to Korsgaard. Rather, he was addressing philosophers who take this sort of impasse to indicate that normative judgments cannot be correct or incorrect. Nevertheless, I believe his observation at least undercuts any assumption that the failure to provide the sort of response I described *must* be problematic.

Nagel's response to the moral skeptic already suggests one way in which our deliberations about our reasons can be guided by a shared standard. Since he thinks it may very well be impossible for us to avoid occupying both the personal and impersonal standpoints, our reasons will be subject to a standard of objectivity that is shared with others who can weigh in on whether our proposed reasons do or do not conform to that standard. While he may not be able to persuade someone who refuses to occupy both standpoints that she *ought* to, Nagel has the means to explain how the rest of us can reason together.

Scanlon also has a story about how we reason together, despite the lack of a knockdown argument that will convince everyone to live morally. Again, he draws a comparison between practical reasons and reasons for belief. In particular, he illustrates his view by describing the ways in which beliefs about set theory have developed over time. As rational beings, we automatically respond to apparent reasons, but sometimes we are mistaken about what our *actual* reasons are. Though he cannot give us a formula for determining what our actual reasons are, he does not think this means we cannot reason about what our actual reasons are. Just as there is no formula for determining in advance what sets are, so there is no formula for determining in advance what our reasons are. Nevertheless, we *can* reason about what sets are.

For years, Scanlon observes, what is now called "Naïve Set Theory" provided the dominant view of sets. According to Naïve Set Theory, for every property, there is a set containing all and only those things that have that property. Naïve Set Theory produces paradoxes which have led logicians to reject it, but Scanlon maintains that Naïve Set Theory might have been rejected simply because it conflicts with the plausible intuition

that the members of a set are prior to the set itself. One of the striking consequences of Naïve Set Theory is that it allows that a set may be a member of itself (consider the set of things that are identical to themselves), which could not be the case if the members of a set are prior to the set itself.[51] Naïve Set Theory has since been replaced with the Iterative Conception of sets, which stipulates that sets are collections of pre-existing elements.[52] Reasoning about sets, Scanlon explains, conforms to a method of Reflective Equilibrium: we begin with intuitions that seem plausible and amend our beliefs to account for conflicts with other plausible intuitions.

Practical reason is like reasoning about sets in that reasons and sets are both abstract objects. Thus, we could have worries about whether we can reason about sets that are similar to the concerns voiced about our ability to reason about reasons. While Scanlon admits that there are significant differences between sets and practical reasons, his depiction of how we reason about sets suggests a way we could reason about reasons. That is, we can reason about both through a process of reflective equilibrium. It should be noted that reflective equilibrium, for Scanlon, plays a different role than it was suggested it might play for Korsgaard. Above, I suggested that achieving a state of reflective equilibrium might make it 'unnecessary' for us to continue asking the normative question, but rejected this maneuver as unsatisfying. Whereas reflective equilibrium in that case is intended to provide *justification* for one's attitudes and beliefs,

[51] Scanlon, "Being Realistic About Reasons," Lecture 4, 6. Some readers might challenge Scanlon's supposition that Naïve Set Theory might have been rejected simply because it was in tension with the intuition that the elements of a set must be prior to the set itself. Whether or not this would have actually happened, insofar as the story Scanlon tells is *plausible*, he provides us with an illustration of a pattern of reasoning we can recognize. That is all he needs to make his point. Thanks to John Giannini for raising this question.

[52] Ibid., 8.

Scanlon is only promoting reflective equilibrium as a method for discovering truth – not as an explanation of the truth of one's beliefs.

We might worry that the results of this method of pursuing reflective equilibrium are not reliable, that we have no reason to assume that our considered judgments accurately reflect what reasons there actually are. But Scanlon thinks this is a fruitless question. He summarizes his position thus:

> The only method we have for arriving at and assessing particular conclusions about reasons is … the method of seeking a reflective equilibrium of our substantive judgments about reasons for action. The question of whether a conclusion we arrive at in this way is correct as a claim about reasons for action or a 'quirk' – a mere manifestation of our particular psychology – is simply and only the first order normative question of whether the consideration in question really is a reason or not. It can only be answered by further reflection of this very same kind.[53]

In other words, the only tool we have for determining what reasons there are is the process of reflective equilibrium, so though we can keep asking ourselves whether a certain consideration really is a reason, we will only be able to answer that question by seeking out reflective equilibrium.

What I hoped to illustrate through Nagel and Scanlon is that realists can offer accounts of how we can reason together. Nagel's condition of objectivity is a standard shared by all (non-solipsist) agents, so others can weigh in on whether a proposed reason conforms to that condition. Scanlon's method of reflective equilibrium likewise invites input from others who may have different intuitions worth our consideration. Given that realists can provide accounts of how we reason together, it is not obviously problematic that realists cannot compel all agents to recognize certain facts as giving them reasons, especially in light of the fact that disagreements with others often end in an impasse.

[53] Ibid., 24-25.

Sometimes we just cannot persuade others to see reasons where we do. As much as we may wish to realize an ideal of rationally compelling others to see the reasons we do, it seems like too much to demand of metaethics.

Conclusion

While my project has been a sustained critique of constructivism's inability to account for deliberate immoral action, any adequate defense of realism must address the constructivist challenge that realism cannot respond to moral skepticism. In this chapter I have attempted to do so in two ways. First, I have attempted to show that the Humean objection that facts (or beliefs) cannot motivate fails. Alternative accounts of reasons are problematic and we have no good reason to accept the Humean objection in the first place. Second, I have addressed what I take to be the correct interpretation of Korsgaard's normative question: can realists show that agents are rationally required to act morally? I have shown that realists can offer responses that mirror Korsgaard's own. That is, realists *can* show that agents are rationally required to act morally in much the way Korsgaard does. We may be tempted to pursue still stronger responses to moral skepticism – responses that will compel agents to act morally, but it is not clear that we should expect to be *able* to offer such a strong response. Thus, I conclude that realists are able to offer perfectly adequate responses to moral skepticism.

CHAPTER FIVE:

Concluding Remarks

I have attempted to argue that moral realism is a stronger metaethical position than constructivism on the grounds that moral realism is able both to explain deliberate immoral action and offer an adequate response to moral skepticism. Constructivism, however, responds to moral skepticism by ruling out the possibility of immoral action. In my first chapter I outlined my basic argument for this conclusion and argued that failure to explain immoral action is problematic. I then suggested that a distinction between objective and subjective reasons is necessary to make room for deliberate immoral action. My second chapter argues that Kant, when faced with a similar challenge, attempts to invoke just this sort of distinction. Whether or not he succeeds, I think it is telling both that he attempts to make this maneuver and that the constructivists who follow in his footsteps cannot. Moral realism, however, is equipped to provide the account we need.

In my third chapter I strengthen my case against constructivism by arguing that in certain cases Korsgaard jeopardizes her response to moral skepticism by allowing for deliberate action that is, strictly speaking, wrong. Constructivism makes the tension between responding to moral skepticism and accounting for deliberate immoral action clear. In the course of explaining how her response to moral skepticism is threatened I also clarify what that response is, setting up my final argument in favor of moral realism in my last chapter. There, I argue that realists are able to offer a response to moral skepticism that is similar in form to the response Korsgaard offers. Furthermore, even if

some are dissatisfied with the strength of realist responses (and Korsgaard's), I contend that these responses are perfectly adequate and all that we could reasonably expect. Thus, realism emerges as the more satisfactory position since it does not force us to choose between responding to moral skepticism and explaining immoral action.

BIBLIOGRAPHY

Allison, Henry. *Kant's Theory of Freedom*. New York: Cambridge UP, 1990.

Anscombe, G. E. M. *Intention*. repr. Cambridge, MA: Harvard UP, 2000.

Aquinas, Thomas. *On Evil*. Translated by Richard Regan. New York: Oxford UP, 2003.

Aristotle. *Nicomachean Ethics*. Translated by Terence Irwin. Indianapolis, IN: Hackett, 1999.

Bagnoli, Carla. *The Stanford Encyclopedia of Philosophy*. Accessed January 14, 2013. http://plato.stanford.edu/entries/constructivism-metaethics/.

Beck, Lewis White. *A Commentary on Kant's* Critique of Practical Reason. Chicago: The University of Chicago Press, 1960.

Brewer, Talbot. *The Retrieval of Ethics*. New York: Oxford UP, 2009.

Cohen, G. A. "Reason, Humanity, and the Moral Law." In *The Sources of Normativity*, edited by Onora O'Neill, 167–188. New York: Cambridge UP, 1996.

Ebels-Duggan, Kyla. "Moral Community: Escaping the Ethical State of Nature." *Philosophers' Imprint* 9, no. 2 (August 2009). http://www.philosophersimprint.org/009008/.

Enoch, David. "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action." *Philosophical Review* no. 2006 (n.d.): 169–198.

Frierson, Patrick. "Two Standpoints and the Problem of Moral Anthropology." In *Kant's Moral Metaphysics*, edited by Benjamin J. Bruxvoort Lipscomb and James Krueger, 83–110. Berlin: Walter de Gruyter, 2010.

Grenberg, Jeanine. "In Search of the Phenomenal Face of Freedom." In *Kant's Moral Metaphysics*, edited by Benjamin J. Bruxvoort Lipscomb and James Krueger, 111–130. Berlin: Walter de Gruyter, 2010.

Kain, Patrick. "Self-Legislation in Kant's Moral Philosophy." *Archiv Für Geschichte Der Philosophie* 86, no. 3 (2004): 257–306.

Kant, Immanuel. *Critique of Practical Reason*. Translated by Mary Gregor. New York: Cambridge UP, 1997.

———. *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor. New York: Cambridge UP, 1998.

———. *Religion Within the Limits of Reason Alone*. Translated by Theodore M. Green and Hoyt H. Hudson. New York: Harper & Row, 1960.

———. *The Metaphysics of Morals*. Translated by Mary Gregor. New York: Cambridge UP, 1996.

Korsgaard, Christine M. "Acting for a Reason." In *The Constitution of Agency*, 207–229. New York: Oxford UP, 2008.

———. "Aristotle and Kant on the Source of Value." In *Creating the Kingdom of Ends*, 225–248. New York: Cambridge UP, 1996.

———. "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations." In *Creating the Kingdom of Ends*, 188–221. New York: Cambridge UP, 1996.

———. "Fellow Creatures: Kantian Ethics and Our Duties to Animals." *Tanner Lectures on Human Values* 24 (February 2004): 77–110.

———. "Kant's Analysis of Obligation: The Argument of *Groundwork I*." In *Creating the Kingdom of Ends*, 43–76. New York: Cambridge UP, 1996.

———. "Kant's Formula of Humanity." In *Creating the Kingdom of Ends*, 106–132. New York: Cambridge UP, 1996.

———. "Kant's Formula of Universal Law." In *Creating the Kingdom of Ends*, 77–105. New York: Cambridge UP, 1996.

———. "Morality as Freedom." In *Creating the Kingdom of Ends*, 159–187. New York: Cambridge UP, 1996.

———. "Realism and Constructivism in Twentieth-Century Moral Philosophy." In *The Constitution of Agency*, 302–326. New York: Oxford UP, 2008.

———. "The Reasons We Can Share: An Attack on the Distinction Between Agent-Relative and Agent-Neutral Values." In *Creating the Kingdom of Ends*, 275–310. New York: Cambridge UP, 1996.

———. "Reply." In *The Sources of Normativity*, edited by Onora O'Neill, 219–258. New York: Cambridge UP, 1996.

———. "The Right to Lie: Kant on Dealing with Evil." In *Creating the Kingdom of Ends*, 133–158. New York: Cambridge UP, 1996.

———. *Self-Constitution: Agency, Identity, and Integrity*. New York: Oxford UP, 2009.

———. "Taking the Law into Our Own Hands: Kant on the Right to Revolution." In *The Constitution of Agency*, 233–262. New York: Oxford UP, 2008.

———. *The Sources of Normativity*. New York: Cambridge UP, 1996.

Mackie, J. L. *Ethics: Inventing Right and Wrong*. New York: Penguin, 1977.

Michalson, Gordon E., Jr. *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*. New York: Cambridge UP, 1990.

Nagel, Thomas. *The Possibility of Altruism*. Princeton: Princeton UP, 1970.

———. *The View from Nowhere*. New York: Oxford UP, 1989.

———. "War and Massacre." In *Mortal Questions*, 53–74. New York: Cambridge UP, 1979.

Pallikkathayill, Japa. "Deriving Morality from Politics: Rethinking the Formula of Humanity." *Ethics* 121 (October 2010): 116–147.

Parfit, Derek. *On What Matters*. Edited by Samuel Scheffler. Vol. I. New York: Oxford UP, 2011.

Potter, Nelson, Jr. "Does Kant Have Two Concepts of Freedom?" In *Akten Des 4. Internationalen Kant-Kongresses*, edited by G. Funke and J. Kopper, 590–596. Berlin: Walter de Gruyter, 1978.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap Press, 1971.

Scanlon, T. M. "Being Realistic About Reasons." presented at the Locke Lectures, Oxford University, 2009. http://www.philosophy.ox.ac.uk/lectures/john_locke_lectures/past_lectures.

———. *What We Owe to Each Other*. Cambridge, MA: Belknap Press, 2000.

Schapiro, Tamar. "Compliance, Complicity and the Nature of Nonideal Conditions." *The Journal of Philosophy* 100 (2003): 329–355.

———. "Kantian Rigorism and Mitigating Circumstances." *Ethics* 117 (2006): 32–57.

Tenenbaum, Sergio, ed. *Desire, Practical Reason, and the Good*. New York: Oxford UP, 2010.

Vogler, Candace. *Reasonably Vicious*. Cambridge, MA: Harvard UP, 2002.

Wallace, R. Jay. “Normativity and the Will.” In *Normativity and the Will*, 71–81. New York: Oxford UP, 2006.

Williams, Bernard. “Internal and External Reasons.” In *Moral Luck*, 101–113. New York: Cambridge UP, 1981.