

## CHAPTER TWO

### KORSGAARD ON SELF-CONCEPTIONS AND PRACTICAL REASONS

#### 2.1 Introduction

My aim in this chapter will be to reconstruct and discuss two arguments presented by Christine Korsgaard in *The Sources of Normativity* concerning the role of self-conceptions in our practical lives.<sup>1</sup> In Section 2.2, I will examine Korsgaard's argument for the view that all of one's normative reasons are grounded in practical identities and discuss a dilemma that this account faces. Then in Section 2.3, I will examine certain components of Korsgaard's moral theory. While Korsgaard's moral views are not of critical importance for my larger project, since her argument for them might require a modification to the account of practical reasons presented in Section 2.2, it is worth examination. The fruit of this discussion will be: (1) an overview of Korsgaard's account of the relationship between normative reasons and self-conceptions, (2) a summary of a serious problem that it runs into, and (3) an overview of Korsgaard's theory of normative reasons.

#### 2.2 Korsgaard on the Relationship Between Self-Conceptions and Normative Reasons

Korsgaard's argument begins with the claim that what is special about rational beings such as ourselves—what distinguishes us from all other animals—is the fact that we are capable of turning our attention upon our own mental activities. We don't simply have desires and perceptions, we are both aware that we have them and we are capable of conceiving of them as desires and perceptions. But “this sets us a problem

---

<sup>1</sup> I will be treating this text, in other words, in isolation from Korsgaard's other writings (including her papers in *Creating the Kingdom of Ends* and other subsequent work). While there appear to be some possible conflicts between the views expressed in *Sources of Normativity* and Korsgaard's other work, I will not concern myself with them in the pages that follow.

no other animal has....For our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and call them into question” (Korsgaard (1): 93).<sup>2</sup> When I find myself desiring a certain object, I am not compelled by this desire to pursue it, but have the ability (and at least occasionally feel the need) to ask whether I *should* pursue this object, or whether the fact that I have this desire makes it the case that I should pursue it. “[T]he impulse doesn’t dominate me and now I have a problem. Shall I act? Is this desire really a *reason* to act? The reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*” (Korsgaard (1): 93). So I have a problem. I need to decide whether this desire is something that I should act on, and the only way of settling this question is to look at considerations that speak for or against acting on this desire. In other words, I need reasons. Accordingly, Step One of Korsgaard’s argument is the following:

**1. Reflective distance from my desires gives me the inescapable need to act on the basis of reasons (if I am to act at all).**

As it turns out, the fact that I am capable of directing my attention on to my own mental states—that I can and do achieve reflective distance from them—places another important requirement on me as well. “The reflective structure of the mind is a source of ‘self-consciousness’ because it forces us to have a *conception* of ourselves” (Korsgaard (1): 100). The point here seems to be that I am “forced” to have some conception of myself as the thing that sometimes achieves reflective distance from my desires. This is so because I cannot conceptualize a situation (to

---

<sup>2</sup> One might, however, question whether these three activities—(a) becoming aware of one’s mental states, (b) distancing oneself from them, and (c) calling them into question—are the product of the same capacity. This is not necessarily to say that the three activities are unrelated; (a) and perhaps (b) are arguably preconditions for (c). In any event, it seems that (c) is what Korsgaard should be most interested in, since it is this ability (and perhaps proclivity) to call mental states into question that is the source of the demand for reasons that she is after.

speak somewhat loosely) in which *I* am over here and my *desires* are over there without conceptualizing the “I” element of this picture to some degree. This argument gives us Korsgaard’s Step Two:

**2. Reflective distance from my desires also forces me to have some self-conception as (at least) the thing that is distinct from my desires.**

To this point we have been told that reflective distance requires me to act for reasons and hold some self-conceptions; the next claim in Korsgaard’s argument tells us that there is an important relationship between these two requirements. To bring this point out, however, we need to look at some things Korsgaard has to say about “expression.” Consider the following passage:

When you deliberate, it is as if there were something over and above all of your desires, something which is *you*, and which *chooses* which desire to act on. This means that the principle or law by which you determine your actions is one that you regard as being expressive of yourself. To identify with such a principle or way of choosing is to be, in St. Paul’s famous phrase, a law to *yourself* (Korsgaard (1): 100).

This notion of “expression” seems to be rather important to Korsgaard’s argument, given her repeated references to it. On the very next page of the text she claims that “Your reasons express your identity, your nature...” (Korsgaard: 101). What precisely does this talk of expression mean?

Let’s consider an example. If we read “principle or law” as referring to a maxim in a Kantian’s sense of the term, Korsgaard appears to be claiming that if I act on the maxim “Help those in need to relieve their suffering,” I regard this maxim as self-expressive. It isn’t entirely clear, however, what this maxim is really expressive of. Two interpretations of Korsgaard’s discussion suggest themselves. With respect

to the maxim in question, it could (on the one hand) be expressive of the fact that I hold the self-conception “Charitable Humanitarian,” or it could (on the other) simply be expressive of the fact that it is *I* who have decided that this policy of helping the needy is a good idea.

On the first reading, I have an antecedent view of myself as a charitable humanitarian that the policy in question falls out of (and so this policy expresses or gives voice to my antecedent view of myself as a charitable humanitarian). On the second reading, I have no antecedent view of myself as a charitable humanitarian (or at the very least, the reason in question doesn’t fall out of this self-conception), but I am the one who decided that helping the needy would be a good idea; the policy would thus be expressive of me as the one who was responsible for weighing the relevant evidence and judging that there was some value in helping the needy. Call the first reading “strong expression” and the second reading “weak expression.” The important distinction between the two is that, on the first reading, I explicitly hold the self-conception “Charitable Humanitarian” and deliberate from it to a reason for action, whereas on the second, I don’t. With these two possible interpretations in hand, we can now state what I will refer to as Korsgaard’s “expression claim” (given the ambiguities in this claim and the fact that it will be superceded by a later interpretation, I will not present this as a formal step in Korsgaard’s argument):

**Korsgaard’s Expression Claim: My normative reasons are expressive of my self-conceptions.**

This claim is meant to follow from, and summarize, the preceding discussion. For we now know that my practical reasons are in some sense “self-expressive,” and, from the

considerations that led to Step Two, we know that I must see them as expressive of me under a given description of myself.

It is important to note that throughout this argument Korsgaard is speaking of normative reasons, and not merely motivating reasons. While one may question whether Korsgaard can ultimately hold on to a distinction between normative and motivating reasons given how she develops her account, on the assumption that she can accommodate a difference between the two, she is quite insistent that it is normative reasons (considerations that actually count in favor of certain actions) that interest her.<sup>3</sup>

Since I am attempting to reconstruct Korsgaard's argument in both a step-wise fashion and in slightly more detail than she herself does, I must at this point begin to introduce a few steps on her behalf, several of which Korsgaard never explicitly lays out herself. To help make clear where the discussion is going, however, it might be useful to call attention to the conclusion of this first phase of Korsgaard's argument:

**Conclusion (Step 6): My normative reasons for action are grounded in my practical identities.**

Note for the time being that Korsgaard defines a practical identity as: "...a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking" (Korsgaard (1): 101). To begin with some textual evidence for this conclusion (more follows as the argument unfolds), we have first Korsgaard's claim that: "...unless you are committed to some conception of your practical identity, you will lose your grip on yourself as having any

---

<sup>3</sup> To give just one example, (which is made in reference to what will be the conclusion of this first component of her argument) Korsgaard insists that "There is a sense in which these obligations [which stem from one's endorsed self-conceptions] are real—not just psychologically but normatively" (Korsgaard (1): 257).

reason to do one thing rather than another” (Korsgaard (1): 121). And second, “Since you are a human you *must* take something to be normative, that is, some conception of practical identity must be normative for you. If you had no normative conception of your identity, you could have no reasons for action” (Korsgaard (1): 123).

Given Korsgaard’s commitment to Step Six, the question at this point is how Korsgaard goes from her expression claim to this much stronger claim (which comes—in terms of its appearance in the text—one paragraph after Korsgaard’s statement of her expression claim). In particular, it appears as though two central moves need to be accounted for if we are to make sense of the transition from the expression claim to this conclusion. First, we need to account for the fact this claim’s talk of “expression” becomes talk of “grounding,” and second, we need to account for the fact that this claim’s talk of “self-conceptions” becomes talk of “practical identities” (which are a subset of self-conceptions). Perhaps the best way of proceeding would be to include a further step that represents a strengthening of “expression” clause of Korsgaard’s expression claim to get us the “grounding” clause that occurs in the conclusion:

### **3. My normative reasons are grounded in my self-conceptions.**

But how might we justify this move from “expression” to “grounding?” Two approaches stand out as viable options, each of which involves committing ourselves to one of the two interpretations of “expression” discussed above.

First, making sense of the respect in which a self-conception grounds reasons for action is relatively easy if one has a strong interpretation of the term “expression” in mind. Recall that a case of strong expression is one in which it is an antecedent self-conception—typically stated in “thick” terms, such as “Liar” or “Friend of

Susan”—that gives rise to one’s reasons for action. In such cases, the thick self-conception is doing a great deal of work; it is, after all, my antecedent view of myself as being a certain sort of thing that *determines* what my reasons for acting are. Accordingly, on this strong reading of the term “expression,” we are licensed in moving from talk of “expression” to talk of “grounding” simply because the fact that A grounds B is what makes it the case that B expresses A.

However, if we go with the weak expression reading, the thin self-conception itself (which would be simply the description: “thing that has decided that  $\phi$ -ing would be a good idea”) doesn’t seem to be what determines my reasons for action; my reasons are determined rather by the considerations that confront my deliberative gaze. For example, it is not the general description of myself as a deliberator that fixes the content of the reasons I have for helping the needy, it is rather the perceived values that stem from helping the needy that fixes the content of these reasons. Nonetheless, we might still be able to justify the transition from the expression claim to Step Four on this weak reading of “expression” if we have a correspondingly weak reading of “grounding” in mind (i.e. one that means something like “is explanatorily or causally pertinent”). For we could then say that my reasons are “grounded” in my thin self-conception insofar as the fact that I am the one who determined that  $\phi$ -ing would be a good idea is explanatorily or causally pertinent to the existence of these reasons (i.e. the reasons wouldn’t exist if I hadn’t drawn certain conclusions in my deliberations).

This would be, however, a rather uninteresting, if not an altogether trivial account of the ways in which my reasons for acting can be “grounded” in my self-conceptions. All this claim really says is that every time I take myself to have a reason for action, my deliberations played a role in my coming to think that I have this reason. The notion of a self-conception isn’t playing much of a role here; in fact, we can eliminate all reference to self-conceptions (as I have in glossing this claim as I just

did) without losing anything. The notion of a self-conception simply isn't pulling its own explanatory weight on such a view. This, it should be noted, is not the case on the strong reading; when an agent decides that he has a reason to act in a certain way *because* he is a certain sort of thing, the self-conception plays an absolutely ineliminable role.<sup>4</sup>

These problems notwithstanding, there is indeed textual evidence for this weak reading. First, the language that Korsgaard uses in the page 100 citation provided above (in which she talks about being a law unto oneself) is suggestive of a weak interpretation of this relationship. One can find further evidence for this weak reading by looking at Korsgaard's account of what a reason for action is. "The normative word 'reason' refers to a kind of reflective success....[to declare that one has a reason to act means] I'm satisfied, I'm happy, I'm committed, you've convinced me, let's go. [It] mean[s] the work of reflection is done" (Korsgaard (1): 94). A reason for Korsgaard, in other words, can be thought of simply as the outcome of my reflective success, where "I" am understood under the description of the "thin" self-conception discussed earlier. Accordingly, one would not be without warrant in suggesting that Korsgaard intended a weak understanding of this grounding relationship.

However, while there is textual evidence in favor of this weak reading, I think there is better evidence for the view that Korsgaard intended a stronger reading of the grounding relationship. Consider first the following passage, which occurs directly after Korsgaard's discussion of the "expressive" element of reasons and the sense in which an agent is a law to herself.

An agent might think of herself as a Citizen of the Kingdom of Ends. Or she might think of herself as someone's friend or lover, or as a member of a family

---

<sup>4</sup> In calling this claim trivial, I am speaking only of its reference to self-conceptions. As will become clear in Chapter 4, there are aspects of this claim (most notably its implication that reasons are a product of our deliberations) that I find hardly trivial.

or an ethnic group or a nation. She might think of herself as the steward of her own interests, and then she will be an egoist. Or she might think of herself as the slave of her passions, and then she will be a wanton (Korsgaard (1): 101).

All of Korsgaard's examples in this passage (and in several other passages of her book) are of what I have been calling "thick" self-conceptions, or self-conceptions that go beyond the thin description "thing that has decided that  $\phi$ -ing would be a good idea." Second, when discussing the actual decision procedure by which reasons are generated, Korsgaard often refers to stretches of deliberation that move from an antecedent self-conception *to* a particular reason for action: "We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves" (Korsgaard (1): 120). On a similar note, she remarks that "When an impulse—say a desire—presents itself to us, we ask whether it could be a reason. We answer that question by seeing whether the maxim of acting on it can be willed as a law by a being with the identity in question" (Korsgaard (1): 113).<sup>5</sup>

Accordingly, when Korsgaard claims that one's reasons express one's self-conceptions, she appears to have in mind what I have referred to as strong expression. The upshot is that we are licensed in moving from the expression claim's talk of "expression" to Step Four's talk of "grounding" simply because the sense of "expression" at stake in this earlier claim is so strong that it can be reasonably thought of as always implying an underlying "grounding" relationship.

It is important to note, however, that Korsgaard's decision to opt for this reading of the expression claim comes at a high price. Our earlier discussion suggested that her expression claim was actually quite plausible as it stood; i.e. if we were allowed to interpret the sense of expression at stake as weak expression. But by

---

<sup>5</sup> It is worth noting that the various phrases Korsgaard uses in the preceding citation to mark one's reflective endorsement of an action represent a number of different attitudes that don't appear to be equivalent. I take it that a certain understanding of what it means to be "be convinced" is what Korsgaard really wants here (one can "be happy" with or "be committed" to an option without reflectively endorsing it).

insisting that all reasons are strongly expressive of one's self-conceptions, Korsgaard has made her position much more implausible.

How So? As both G. A. Cohen and Thomas Nagel point out in their replies to Korsgaard, it seems very odd (if not patently false) to claim that all of one's reasons for action are grounded in antecedent, "thick" self-conceptions. Cohen, for example, remarks that the claim that one's reasons are grounded in one's self-conceptions simply doesn't ring true,

...except in the trivial sense that, if I treat something as a reason, then it follows that I regard myself as, identify myself as, the sort of person who is treating an item, here and now, as a reason. I do not see that I must consult an independent conception of my identity to determine whether a possible spring of action is to be endorsed or not, nor even that such endorsement must issue in such a conception, other than in the indicated trivial sense (Korsgaard (1): 185).

In effect, Cohen is challenging Korsgaard's blanket claim that "We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves" (Korsgaard (1): 120). While I can only speak for myself and those I have put the question to, my own experience has shown that people only rarely deliberate about what to do by reflecting on what someone with a given self-conception would do in certain situations, such that we could say that this was the aim of their deliberation. Yesterday I took myself to have reason to perform a large variety of actions—making breakfast, preparing lecture notes for my students, reflecting on Korsgaard's account of practical reasons—and in none of these cases did any explicit consideration of self-conceptions come up. In short, it is simply implausible to claim that one does (or perhaps should) deliberate about what to do in all cases by consulting antecedent, thick self-conceptions, which is precisely what Korsgaard suggests.

I take Nagel to be making a similar point when he remarks:

To decide from the reflective standpoint what to do you have eventually to stop thinking about yourself and think instead about the question at issue—not in this case about what entities the world contains, but about whether what has made you want to do something is really a reason to do it. The answers to such questions may partly determine your identity, but they don't derive from it (Korsgaard (1): 206).

So what is going on here? Why does Korsgaard think that her Step Three claim is correct? Or to put the point in terms of the distinctions I have been drawing, why does she think that the only legitimate or interesting form of expression is strong expression? As far as I can tell, Korsgaard offers no real argument for this claim. (It is interesting to note that in the reply section of her book, Korsgaard doesn't address Cohen's and Nagel's objections on this point).

To sum up, we can pose the central problem in the form of a dilemma, and moreover, one that can be specified in two different ways: one that is specific to the expression claim to Step Four transition of Korsgaard's argument, and one that is more general to the expression claim itself. The first specification of this dilemma is as follows. On the one hand, Korsgaard could have intended a relatively liberal interpretation of her expression claim, whereby it could be claimed that one's reasons "express" one's self-conception in the weak sense of this term. This move would have the benefit of making the expression claim plausible, but at the expense of invalidating the move to Step Three (since weak expression cannot be glossed as representing a grounding relationship in any interesting sense of the term). Alternatively, Korsgaard could (and I believe ultimately does) insist that she has only a strong interpretation of "expression" in mind in her expression claim. This has the benefit of validating the move to Step Four, but at the expense of making the expression claim implausible.

The second (and perhaps more straightforward) way of specifying this dilemma puts the issue squarely in terms of Korsgaard's expression claim. For on the basis of our two interpretations of this Thesis, we can say *either* that this claim can be made plausible at the expense of making it trivial by going with a weak interpretation of "expression," *or* it can be made non-trivial at the expense of making it implausible by going with a strong interpretation of this term.

It is important to note that Korsgaard faces this dilemma (as well as some other problems that will be discussed in Chapter Five) because of her insistence that the relationship between self-conceptions and practical reasons is both (a) universal (i.e. holds in all cases), and (b) unidirectional (self-conceptions rise to practical reasons in all cases, but not vice versa). As it turns out, we needn't abandon both claims to avoid this dilemma; the problem at issue is the result of their conjunction. As I will be discussing in later chapters, if we abandon claim (b), we can hold on to claim (a) and avoid this dilemma, provided that we make a few other modifications to our account. In terms of lessons learned, this diagnosis of Korsgaard's dilemma suggests that a promising strategy for developing this relationship will involve positing a bidirectional account of it, whereby practical reasons and self-conceptions interact more dynamically.

Putting this dilemma off to the side for the moment (we will return to it in later chapters), our next task is to make sense of Korsgaard's move from talk of "self-conceptions" in her expression claim to talk of "practical identities" in her conclusion. The first thing to notice is that this move does constitute a significant strengthening of the position in question, insofar as practical identities represent a fairly small subset of self-conceptions. We have seen why Korsgaard thinks that reasons must be grounded in self-conceptions, but why must they be grounded in practical identities (i.e. self-conceptions that make the agent's life and actions "worthwhile" from her own point of

view)? The answer, I think, turns on Korsgaard's commitment to a specific version of reasons internalism and the related thesis that all normative reasons are grounded in the antecedent desires of agents. Before actually making a case for this claim, however, let us lay out Step Four of the argument as follows:

**4. My normative reasons are grounded in objects of my concern (things that I desire or care about).**

This is a claim that Korsgaard doesn't really argue for in *Sources of Normativity*, but it is clearly one that she is committed to. We have already seen some support for this claim in the citations presented earlier in support of Korsgaard's conclusion to this argument, which holds that our normative reasons for action come from our practical identities (which are again self-descriptions that we care about). That all of our normative reasons are grounded in cared-for self-descriptions is simply a more narrow specification of the more general claim presented in Step Four.

This view also figures prominently in Korsgaard's discussion of the normative reasons that are tied to our moral obligations, which serve as the focus of much of her book. Korsgaard claims repeatedly that we are only obligated to perform actions that we have reflectively endorsed. This claim, together with Korsgaard's insistence that reflective endorsement is closely linked to motivation, implies a close connection between being obligated to  $\phi$  and being motivated to  $\phi$ . This link between obligation and motivating reflective endorsement, after all, is supposed to be what accounts for the "normativity" of obligations on Korsgaard's view. When I am motivated to execute my obligation to  $\phi$ , this will be because I already care about something (i.e. a practical identity) that gives rise to my obligation to  $\phi$ .

For example, Korsgaard claims that one's obligation to comply with the rule: "A psychiatrist doesn't violate the confidence of her patients" isn't mysterious at all. "No 'ought' is needed here because the normativity is built right into the role" (Korsgaard (1): 101). For Korsgaard an "ought" would be superfluous in such a case because the person in question *already* cares about and endorses the self-conception "Psychiatrist," and hence she is already disposed to act in response to obligations that spring from it. On the next page of the text she goes on to argue:

It is the conceptions of ourselves that are most important to us that give rise to unconditional obligations. For to violate them is to lose your integrity and so your identity, and to no longer be who you are. That is, it is to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living and your actions to be worth undertaking. It is to be for all practical purposes dead or worse than dead (Korsgaard (1): 102).

The central but unstated question I take Korsgaard to be posing here is: if normative reasons cannot be grounded in objects of such great importance to an agent (such as highly valued self-conceptions), what else could possibly ground them?

To further flesh out this view of obligation and its sources, Korsgaard goes on to remark that "...in one sense, the obligatory is like the visible: it depends on how much of the light of reflection is on" (Korsgaard (1): 257). This suggests that the only obligations that we are really under—and the only normative reasons that we have for acting on a given obligation—are ones that we acknowledge. If I remain unconvinced that I have normative reason to  $\phi$ —which, according to Korsgaard, typically happens when  $\phi$ -ing is unrelated to things that I care about—I simply don't have a normative reason to  $\phi$ . Korsgaard appears to strengthen this claim at one point by suggesting that it is only things that we *actually* care about—and not things that we might hypothetically care about or rationally ought to care about—that give us normative

reasons. "...[I]t is the endorsement, not the explanations and arguments that provide the material for the endorsement, that does the normative work" (Korsgaard (1): 257).

Finally, it is worth mentioning that Korsgaard's theory of value also underwrites her commitment to Step Four. Given the generally accepted view that values and normative reasons are closely associated, we would expect that a view that links normative reasons to an agent's desires would also link values to an agent's desires as well. And this is precisely what we find in Korsgaard. She claims that "Good maxims are intrinsically normative entities, but they are also the products of our own legislative wills. In that sense, values are created by human beings....the maxim isn't a law until we will it, and in that sense create the resulting values" (Korsgaard (1): 112). Since Korsgaard's value constructivism ties value to one's willed maxims (which Korsgaard often describes as proposals that we act on particular desires), we shouldn't be surprised to find her claiming that normative reasons are tied to one's desires as well. For Korsgaard, this is what reflective endorsement and legislative willing is all about. We construct values by endorsing maxims that recommend acting on the basis of certain desires. (Recall that this entire argument kicked off with the idea that agents need to figure out which of their desires to act upon).

As we will see in the next section, this may not be Korsgaard's last word on the role that desires play in grounding our normative reasons. This picture is complicated by some further remarks Korsgaard makes about possible moral obligations to act in certain ways. For the time being, it will suffice to note how much textual evidence there is to support attributing Step Four to Korsgaard.

This brings us to our Step Five claim, which consists merely in a stipulative definition.

The conception of one's identity in question here is not a theoretical one, a view about what as a matter of inescapable scientific fact you are. It is better understood as a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking. So I will call this a conception of your practical identity (Korsgaard (1): 101).

From this passage we get the Stage Five claim, which is as follows:

**5. A practical identity is a self-conception that the agent cares about.**

We are now in a position to see how Korsgaard can move from her expression claim to her conclusion. Steps Three, Four, and Five jointly entail Korsgaard's conclusion, which is again the claim that my reasons for action are grounded in my practical identities. For if my reasons for action must be grounded in both (a) my self-conceptions, and (b) objects that I care about (such that (a) and (b) are two necessary conditions for "reasonhood"), then we can conclude that my normative reasons must be grounded in self-conceptions that I care about, which are just practical identities.

Before moving on, it might be useful to reflect on just how strong this conclusion is. Korsgaard is not arguing that some of our normative reasons stem from our practical identities, or that most of them do, or even simply that the strongest reasons we face are grounded in our practical identities. She is claiming that every normative reason I have for acting is grounded in a description of myself under which my life has meaning.

### **2.3 Moral Identity and Korsgaardian Normative Reasons Revisited**

The next phase of Korsgaard's project consists of an argument for a particular moral view. While many of the details of this argument aren't relevant to the main themes of my project, we must look at certain aspects of her moral argument quite

closely because they complicate the story told to this point. The view of normative reasons that I've been presenting on Korsgaard's behalf has been relatively straightforward; according to it, my normative reasons are grounded in my practical identities. However, further claims made by Korsgaard in later stages of her book throw this view into question. As I'll soon discuss, Korsgaard at least flirts with the view that I might have normative reason to act morally regardless of whether I endorse self-conceptions with moral implications.

The best way to capture Korsgaard's argument for morality is to present her view as a series of first person reflections (my reasons for proceeding in this way will become clear shortly). Accordingly, let us begin by considering the following passage, which sums up the central steps of Korsgaard's moral argument that will concern us.

...[T]hese conceptions [i.e. practical identities] are contingent, [and] one or another of them may be shed....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. But this reason for conforming to your particular 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 (Korsgaard (1): 120-1).

Let's consider this argument step by step. Given Korsgaard's Step Six claim, I can see that if all of my reasons are grounded in my practical identities, then I can have no reason to do anything (including, presumably, reason to go on living or even to end my life) unless I have some practical identity that generates reasons for me. Based on our Step One claim, I can also see that this fact is an important one. The

reflective structure of my consciousness presents me with a problem: I *need* reasons for acting one way rather than another. The implication of these two steps is the following:

**7. I need to have reason-grounding practical identities.**

While the pattern of inference underlying this bit of reasoning is fairly straightforward—realizing that there is only one solution to a serious problem leads me to see that I need to adopt this solution—it plays a central role in Korsgaard’s larger project. This pattern of inference, presented in the context of John Rawls’ concept/conception distinction, proves to be the basis of Korsgaard’s general view of normativity. “The normative force of the [solution] is established in this way. If you recognize the problem to be yours, and the solution to be the best one, then the solution is binding upon you” (Korsgaard (1): 114). Considerations are normative for us (some sort of “ought” attaches to them), because they represent solutions to problems that we already worry about, or perhaps cannot help but worry about. The particular problem in this case is that I need reasons for acting. The solution is that reasons spring from my practical identities. Hence, this solution is binding upon me; I face a demand that I acquire or maintain practical identities since they generate reasons, which solves my problem.

Recognizing this need of mine is in Korsgaard’s language to recognize that I have a “moral identity” (a term that Korsgaard uses almost interchangeably with the term “humanity”).<sup>6</sup> The next step in the argument is accordingly nothing more than a stipulative definition.

---

<sup>6</sup> It is important to note the structure of Korsgaard’s argument at this point in the dialectic. I don’t take Korsgaard to be claiming that we are justified in acting in accordance with a *given* practical identity simply because it is *a* practical identity (and hence *a* solution to the problem we face). While it is true that we need at least some practical identities, the particular ones we hold are not justified wholly by

**8. “Moral identity” refers to one’s identity as a being who needs reason-grounding practical identities.**

Swapping this terminology into our Step Seven claim, we can rephrase this newly discovered need as follows:

**9. I need to act in accordance with my moral identity.**

At this stage in the argument, things become more complicated. The next inference that I (as a reader trying to understand the predicament I am in) am supposed to make according to Korsgaard, is that this latest discovery of mine (i.e. Step Nine) implies the following:

**10. Given the truth of Step Nine, I ought to value my moral identity.**

Why exactly Korsgaard thinks that this is a natural conclusion for me to draw is unclear, but that she thinks I should is not. Consider the following citations.

---

this fact of “identity necessity,” as Samuel Kerstein has suggested (Kerstein: 49). If this was Korsgaard’s view, Kerstein would be right to object to it, but I think Korsgaard introduces the notion of “identity necessity” (as Kerstein calls it) not as a means of justifying one in holding a *given* practical identity, but rather as a premise in her larger argument for the claim that one ought to value one’s moral identity. Korsgaard is no doubt partially responsible for this confusion given the way that she words many of her claims. A more promising line of objection to this phase of Korsgaard’s argument is offered by Christopher Gowans, who suggests that Korsgaard’s general account of normativity—which on Gowans’ view finds its seat in the will’s (universalizable) end-setting activity—is incapable of accounting for our decisions to adopt particular practical identities over others. See, for example (Gowans: 557). I am on the whole quite sympathetic to Gowan’s arguments on this point, but won’t be addressing them here.

- “...[I]f you value anything at all, or if you acknowledge the existence of any practical reasons, then you must value your humanity as an end in itself” (Korsgaard (1): 125).
- “Moral identity also stands in a special relationship to our other identities. First, moral identity is what makes it necessary to have other forms of practical identity, and they derive part of their importance, and so part of their normativity, from it (Korsgaard (1): 129).”
- “Since you cannot act without reasons and your humanity is the source of your reasons, you must value your own humanity if you are to act at all” (Korsgaard (1): 123).

The general idea seems to be that, if I reflect sufficiently, I will see that I ought to value my moral identity if I take myself to have any reasons whatsoever. Why does Korsgaard think this? While a definite answer isn't forthcoming, it is possible that Korsgaard is relying on some other, more general claim about the nature of reasons for action that links them intimately to values. She might think, for example, that the fact that I have reason to  $\phi$  entails that  $\phi$ -ing, or some result of  $\phi$ -ing, is valuable (or is at least worth treating as valuable). This wouldn't be an entirely absurd suggestion to make, but it isn't clear that it is true.

Rachel Cohon, for example, has wondered why we must value that aspect of us (namely our reflective consciousness) that is responsible for creating the problem that practical identities are supposed to solve in the first place. In “The Roots of Reasons,” she makes the following point:

But how does finding a particular identity and endorsing it commit [the agent] to endorsing the reflectivity that put him into this bind in the first place?....Why couldn't he think of his reflective nature as a weakness, imposing a need for identification rather like an addict's need for her drug or a diabetic's need for

insulin?...I do not see that he must value his status as one who has this need, rather than regard the need as a weakness to cope with as best he can (Cohon: 81).

In other words, the most that Korsgaard should be able to conclude is that, insofar as I take my need for practical identities seriously, I am committed to taking my moral identity seriously as well; but “taking seriously” and “valuing” are two very different things.

It isn't clear, in other words, that Korsgaard's argument is valid. She claims at one point that “Since you cannot act without reasons and your humanity is the source of your reasons, you must value your own humanity if you are to act at all” (Korsgaard (1): 123). But is it the case that if I value B, and A is the source of B, I must value A as well? In many instances this might be the case. For example, not only do I value the coffee that comes out of my coffeemaker, but I value my coffeemaker as well as the source of great coffee. In other cases, however, this claim is less plausible. I may greatly value the \$5,000 that a loanshark lends me at a 40% monthly interest rate to pay off my gambling debts, but I cannot see how I am at all committed to (truly and sincerely) valuing the loanshark. There may be a sense in which I am thankful that he was there to lend me money when I needed it, but I may be thankful in this sense while thinking that the world would be a much better place without people like him in it.

Setting aside problems of this sort, the next step in the argument is simply the following:

**11. I ought to value myself as a being with moral identity.**

This is not the end of Korsgaard's argument for morality. The next few steps—which gloss over some rather important discussion in the text that I will not discuss here since it isn't relevant to present aims—look something like the following:

**12. I ought to value others as beings with moral identities as well.**

**13. Kantian morality falls out of valuing moral identity in myself and in others.**

**14. I ought to abide by Kantian morality.**

Let us assume for the moment that plausible inferences link Step 11 to Step 14—inferences that may be in the offing, even if Korsgaard's quasi-Wittgensteinian route doesn't work. The important point given our interests is that if this reconstruction of Korsgaard's view is accurate, it may not be the case that the only normative reasons that I have for acting are tied to the practical identities that I actually hold. Perhaps I have normative reason to act morally even if I don't acknowledge the value of my moral identity, for example.

Is this Korsgaard's view? This, it turns out, is a very difficult question. The story I've been telling about how morality can get its purchase on agents is fairly straight-forward if we focus on agents who have thought about Korsgaard's arguments and have come to endorse their moral identity. If these are the only sorts of cases that we look at, no modifications to our initial, Section 2.2 account need to be made to accommodate moral normative reasons (since all of the agents in question endorse their moral identity as a practical identity, and so all the normative reasons we are looking at are indeed tied to practical identities).

However, things get very complicated when we ask what we should say about agents who either don't consider this second phase of the argument or are not

convinced by it. The question again is whether such agents have normative reason to act morally despite the fact that they don't endorse their moral identity. Based on the first part of the argument canvassed in the previous section, the answer would seem to be "no;" Korsgaard has claimed that all of one's normative reasons for action are grounded in endorsed self-conceptions. Based on this second phase of the argument, however, the answer isn't at all clear. Since the second phase of the argument is presented as a story that an agent might tell himself (this is how Korsgaard presents it), it is hard to say whether she (a) simply thinks that most reflective people will come to endorse their moral identity (and will *therefore* have reason to act morally), or (b) thinks that everyone has these reasons regardless of whether they have reflected in the right ways.

To help make sense of these two possibilities, consider the following two claims about the grounds of one's normative reasons:

- (A) My normative reasons are grounded in my practical identities (endorsed self-descriptions).
- (B) My normative reasons are grounded in both my practical identities and descriptions of me that are entailed by these practical identities (like my moral identity).

To rephrase our question, we need to know whether the first-person walkthrough of the second phase of Korsgaard's argument implies (B) or not. It is worth noting that there is textual support for (B). Consider the following four citations (the emphasis is mine throughout):

- "Moral identity and the obligations it carries with it are therefore *inescapable and pervasive*....moral identity is *necessary*" (Korsgaard (1): 121-2).

- “You *are* an animal of the sort I have just described. And that is not merely a contingent conception of your identity, which you have constructed or chosen for yourself, or could conceivably reject. *It is simply the truth*” (Korsgaard (1): 123).
- “Moral identity is therefore *inescapable*” (Korsgaard (1): 129-30).
- “Practical conceptions of your identity which are fundamentally inconsistent with the value of humanity *must* be given up” (Korsgaard (1): 130).

These citations are full of talk of necessity, what is inescapable, and “truth.” Unlike her claim that when it comes to practical identities, no question of their truth or falsity is at issue (the issue is rather their endorsement), here Korsgaard suggests that when it comes to the existence of this Step 10 and 11 “ought,” no question of endorsement is at issue (the issue is rather the truth of one’s moral identity).

Based on these sorts of claims, one charitable interpretation of Korsgaard’s final position is precisely claim (B) above; her view is that your endorsement of any practical identity rationally commits you to the endorsement of your moral identity, which in turn has moral implications regardless of whether you are aware of them. This conclusion is also suggested by one of Korsgaard’s remarks about agents that haven’t reflected in the ways she recommends. She claims that this sort of person’s “...obligation to be a good person is therefore deeper than his obligation to stick to [an immoral code underwritten by explicitly endorsed practical identities]” (Korsgaard (1) 258).

So why not simply attribute (B) to Korsgaard and say that her developed view doesn’t emerge until later sections of the book? The problem with this suggestion is that a majority of Korsgaard’s remarks—to include remarks occurring on the penultimate page of the book—seem to support the claim (A) view that it is only the practical identities that one actually holds that give one normative reasons. Recall the

textual claims rehearsed in Section 2.2 in support of Step Four of Korsgaard's argument. There she compared obligation to the visible, suggesting that only obligations that an agent acknowledges are genuine. She also claimed that it was acts of actual endorsement, and not arguments favoring certain endorsements, that do the "normative work."

A close examination of the way that Korsgaard phrases many of her remarks also count in favor of (A). After recapping her argument for moral identity and Step 14, Korsgaard claims: "I take *this* argument to show that any reflective agent *can be led to acknowledge* that she has moral obligations" (Korsgaard (1): 125, my emphasis). It is not unimportant that Korsgaard doesn't simply say: "I take this argument to show that any reflective agent *has* moral obligations;" her claim is rather that any reflective agent can be shown that she has moral requirements, which is what Korsgaard really seems to be after most of the time.<sup>7</sup> Given these sorts of remarks, it doesn't seem as though we can easily declare that (B) is Korsgaard's considered position.

So what is going on here? Based on both this ambiguity and some other remarks that Korsgaard makes, part of the problem appears to be that Korsgaard is sensitive to two conflicting pressures that aren't easily reconcilable. The first, which we might call an "internalism pressure," involves the idea that a consideration cannot really count as a normative reason if an agent is unmoved by it, even upon reflection. Korsgaard's sensitivity to this pressure is especially pronounced in her objections to moral realism. According to Korsgaard, substantive moral realism runs off the rails by insisting that agents have normative reasons to act in ways that they don't recognize or endorse. How are we supposed to accept a moral theory, she wonders, that posits

---

<sup>7</sup> To give one more piece of textual evidence for this claim, recall the p. 123 citation discussed earlier: "It is because we are [reflective, reason-seeking] animals that our practical identities are normative for us, and, *once you see this*, you must take this more fundamental identity, being such an animal to be normative as well (Korsgaard (1): 123 my emphasis).

normative reasons that some agents cannot become motivated to act upon? This general worry raises similar problems for many accounts of non-moral reasons. This internalism pressure recommends against any view that posits a gap between what I can be motivated to do and what I have normative reason to do.

The other pressure Korsgaard is sensitive to is a general “objective moral pressure,” which pushes her toward the view that agents can be subject to moral requirements even if they aren’t aware of them. As a proponent of moral realism (albeit of a “procedural” sort), Korsgaard is (usually) bothered by the notion that a person can avoid moral obligations by simply refusing to acknowledge them. Surely it is morally wrong to murder innocent people—and agents accordingly ought not to do so—regardless of what a potential murderer happens to think.<sup>8</sup>

Each of these two pressures is most at home with one of the two views of normative reasons—(A) and (B)—listed above (the internalism pressure pairs with (A), and the objective moral pressure pairs with (B)). A wholehearted submittal to either pressure counts in favor of adopting the associated thesis of normative reasons. The problem, of course, is that these two pressures urge conflicting views. To put this conflict in the form of a dilemma, Korsgaard has to accept one of the following options. Either (a) there can be a gap between what I can be motivated to do and what I have normative reason to do, or (b) I can have an overriding normative reason to act immorally. Korsgaard is apparently quite bothered by both possibilities.

The final three pages of *The Sources of Normativity* serve as an excellent illustration of Korsgaard’s inability to ally herself decisively with one of these pressures against the other. In the discussion in question, Korsgaard considers the question that we examined above: what should we say to someone who hasn’t

---

<sup>8</sup> My thinking on this tension has benefited greatly from discussions with William Fitzpatrick. For a very interesting alternative assessment of Korsgaard’s bind, see his “The Practical Turn in Ethical Theory: Korsgaard’s Constructivism, Realism, and the Nature of Normativity.”

reflectively come to see that he ought to endorse his moral identity? The case study Korsgaard adopts to explain her answer involves the case of a Mafioso who thinks he has normative reason to do some pretty awful things.

Korsgaard begins this discussion by claiming that, since the Mafioso has endorsed the practical identity “Mafioso,” he has perfectly good normative reasons to kill people, intimidate them, etc. She claims:

It would be intellectually tidy, and no doubt spare me trouble from critics, if I now said that only those obligations consistent with morality are ‘real’ or...‘genuine.’ Then I could say....that there’s no obligation here [to live up to the Mafioso code], only the sense of obligation: no normativity, only the psychic appearance of it....But I am not comfortable with this easy way out...there is a real sense in which you are bound by a law you make for yourself until you make another (Korsgaard (1): 257).

Shortly after, Korsgaard makes the remark mentioned earlier about it being actual endorsement, and not the arguments that count in favor of endorsing a particular proposition, that do the normative work. These claims all seem to point toward thesis (A) about normative reasons; I have normative reason to do only those things that are consistent with my practical identities.

However, closer inspection reveals that these remarks, while certainly suggestive of thesis (A), don’t absolutely require it. Korsgaard could hold, after all, that the Mafioso has conflicting obligations: an obligation to perform immoral acts as a Mafioso, and an obligation to act morally. Indeed, Korsgaard remarks in the next paragraph that she doesn’t think that “...all obligations are moral, or that obligations can never conflict” (Korsgaard (1): 157). On the basis of this remark alone, it seems as though Korsgaard is now paving the way for thesis (B) and the notion that normative reasons *needn’t* come from practical identities.

Later on in this same paragraph, however, we come across Korsgaard's claim that "if one holds the view, as I do, that obligations exist in the first-person perspective, then in one sense the obligatory is like the visible: it depends on how much of the light of reflection is on" (Korsgaard (1): 257). If we take this claim seriously, Korsgaard has once again apparently shifted the drift of her discussion back towards thesis (A) and the view that unacknowledged obligations (and normative reasons) aren't really obligations (or normative reasons) at all.

If this wasn't confusing enough, in the next few sentences we get another shift back towards thesis (B). Korsgaard first claims that she doesn't "...mean to suggest that the Mafioso's obligation to give up his immoral role is something that exists only in the perspective of the rest of us, and not in his own" (Korsgaard (1): 257). She then remarks that "his obligation to be a good person is therefore *deeper* than his obligation to stick to his code" (Korsgaard (1): 258). The closing note of her discussion thus swings the pendulum back towards (B). Given this many shifts of apparent allegiance between (A) and (B) occurring in the span of one and a half pages, it is indeed difficult to figure out what Korsgaard's considered view turns out to be.

Is there a way to make sense of Korsgaard's closing discussion? The only way to render the Mafioso passages coherent would involve claiming that, while there is a sense in which the Mafioso ought to act morally, there is another sense in which it is perfectly acceptable for him to act immorally. Underwriting this view would be a third, more complicated thesis, which looks something like this:

(C) While it is true that my normative reasons for acting are only grounded in my practical identities (the only thing I "ought" to do is act in accordance with them), it is also the case that I "ought" to endorse my moral identity (and thereafter recognize genuine normative reasons to act morally).

On this view, I “ought” to act in accordance with my practical identities, and yet I also “ought” to endorse a particular self-conception and recognize it as reason-giving. The first clause is supposed to speak to the internalism pressure, and the second is supposed to speak to the objective moral pressure.

However, reflection reveals that while this sort of view might render the Mafioso discussion less incoherent (since Korsgaard’s various remarks about what the Mafioso “ought” to do are no longer in *prima facie* conflict), it does so only at the expense of committing what appears to be an even greater incoherence. In order to make sense of the various Mafioso claims that Korsgaard makes, we have to understand (C) as positing a “dichotomy of oughts.” In other words, while (C) claims both that I ought to treat my practical identities as reason-giving, and that I ought to treat my moral identity as a practical identity, we must understand (C) in a way that blocks any inference from the second “ought” to the first “ought.” The force of the theoretical ought covered by the second clause and the force of the practical ought covered by the first must be non-transitive.

To bring this point out, consider the following argument:

- (1) You are a being with moral identity (this is simply a fact about you).
- (2) You ought to hold the self-conception “Being With Moral Identity.”
- (3) Given some further premises, it turns out that agents who hold the self-conception “Being With Moral Identity” ought to abide by Kantian morality.
- (4) You ought to abide by Kantian morality (i.e. you have normative reason to act morally).

Setting (C) aside for a moment, it would be relatively uncontroversial to claim that (4) follows quickly and easily from (1) - (3). Since beings with moral identity ought to view themselves as such, and beings that view themselves as such ought to act

morally, you (as a being of this sort) ought to act morally. One “ought” leads to another, which allows us to move with ease from (1) to (4).

However, it appears as though Korsgaard wants to block any inference from claims (1) - (3) on the one hand, to claim (4) on the other (or between one “ought” and the other). The upshot is that while Korsgaard allows that claims (1) through (3) all rightly apply to the Mafioso, she won’t allow us to infer that he has normative reasons to act morally. (If she did allow this inference, Korsgaard’s discussion of the Mafioso wouldn’t contain the ambiguities and apparent flip-flops discussed above...she would simply say that the Mafioso has normative reason to act morally, and that would be that).

This, however, is a somewhat odd view to hold. Looking at the argument for (1) to (4) above, the structure of the argument appears to be valid, but Korsgaard apparently doesn’t think that it is. She doesn’t think that non-practical oughts entail practical oughts in the relevant way, even with suitable linking premises. The upshot of all of this is that thesis (C) might be able to render Korsgaard’s discussion of the Mafioso coherent, but only at the expense of positing a rather bizarre “dichotomy of oughts” that has counterintuitive implications.

Perhaps even worse, it isn’t clear that this sort of view would fully satisfy the objective moral pressure to which Korsgaard is sensitive. According to (C), while Korsgaard gets to say that the Mafioso “ought” to abide by Kantian morality (in some rather weak sense), given the uncertainty of what exactly this means (it apparently *doesn’t* mean that he has normative reason to act morally), one could object that the “ought” in question is toothless. As far as I can tell, the most that this view allows us to say is that while the Mafioso has an all-things-considered normative reason to act immorally, we are allowed to criticize him for not being sufficiently reflective.

Speaking as someone who feels the gentle push of what I've been calling the objective moral pressure, I for one don't find that this account properly answers to it at all.

While it would be nice to conclude this discussion with something a bit more definitive than this sketchy reconstruction, luckily a clear cut understanding of Korsgaard's view on moral requirements isn't necessary for the discussion that follows. The important point is that, whatever the specifics of Korsgaard's account turn out to be, she apparently wants to stop short of claiming that all agents have normative reason to act morally. I will accordingly proceed in what follows as though Korsgaard is committed to the view that all of one's normative reasons are grounded in practical identities, and will set aside questions of how people might nonetheless be under some sort of pressure to reflect in ways that would lead them to act morally. This specific view of Korsgaard's—together with the problems that stem from her commitment to a unidirectional account of the relationship between reasons and self-conceptions—are what will serve as our focus in upcoming chapters.