

Michael Bukoski

Korsgaard's Arguments for the Value of Humanity

[forthcoming in *Philosophical Review*; please cite published version]

ABSTRACT: In *The Sources of Normativity* and elsewhere, Korsgaard defends a Kantian ethical theory by arguing that valuing anything commits one to valuing humanity as the source of all value. I reconstruct Korsgaard's influential argument to show how she can resist many of the objections that critics have raised. I also show how the argument fails because, at a crucial point, it begs the question in favor of the value of humanity. It thus fails for internal reasons that do not depend on rejecting Korsgaard's metaethical constructivism or her conception of rational agency.

In *The Sources of Normativity*, Korsgaard (1996c) argues that anyone who values anything is committed to valuing humanity as the source of all value. Following Kant, Korsgaard (1996a, 110-14) identifies humanity, or rational nature, with the capacity to set ends freely through the exercise of practical reason rather than following ends given by instinct. She (1996c, 92) postpones a detailed account of what valuing humanity requires, but the general idea is that we must harmonize our ends with those of others, attaching normative weight to their ends as well as to our own. "Enlightenment morality is true," she concludes. "Human beings are valuable" (1996c, 123). Moreover, anyone can recognize this truth upon sufficient reflection: "any reflective agent can be led to acknowledge that she has moral obligations" (1996c, 125).

Korsgaard offers two arguments for this striking claim. One of them she attributes to Kant; call it *Kant's Argument*. The other, *Korsgaard's Argument*, she describes as "just a fancy

I would like to thank Tom Christiano, David Enoch, Connie Rosati, Mark Timmons, and two anonymous referees for very helpful comments and suggestions.

new model” of Kant’s Argument; the latter presents the same argument as Korsgaard’s Argument but “in a much simpler form” (1996c, 122).¹ Perhaps as a result, some critics (for example, Regan 2002) have rejected Korsgaard’s arguments for the value of humanity after consideration of Kant’s Argument alone. If it were as similar to Korsgaard’s Argument as she suggests, then this might be no problem. But Korsgaard undersells her accomplishment: her version of the argument provides a novel and powerful defense of one of the most controversial premises of its predecessor. Thus, I have two aims in this article. The first is to highlight a serious problem with Kant’s Argument and to explain how Korsgaard’s Argument seems to avoid it. The second is to show that, despite its promise, Korsgaard’s Argument ultimately begs the question in favor of the value of humanity. It thus fails in a way that even those in agreement with Korsgaard’s general metaethical approach and conception of rational agency should acknowledge. Having stalked realists and skeptics alike for more than twenty years, the argument deserves a decisive rest.

1. Kant’s Argument

Korsgaard’s most extended discussion of what I am calling Kant’s Argument is in “Kant’s Formula of Humanity” (1996a). But she summarizes and at least largely endorses that argument in *The Sources of Normativity* (1996c), so I will begin there and refer back to the earlier essay as necessary. Korsgaard writes:

¹ For critical discussion of what I am calling Kant’s Argument, see, for example, Gaut 1997, Gibbard 1999, Kerstein 2001, Regan 2002, FitzPatrick 2005, and Langton 2007. For critical discussion of Korsgaard’s Argument, see, for example, Cohon 2000, Kerstein 2001, Gowans 2002, FitzPatrick 2005, and Street 2012.

[Kant] started from the fact that when we make a choice we must regard its object as good. . . . He asked what it is that makes these objects good, and, rejecting one form of realism, he decided that the goodness was not in the objects themselves. Were it not for our desires and inclinations . . . we would not find their objects good. Kant saw that we take things to be important because they are important to us—and he concluded that we must therefore take ourselves to be important. In this way, the value of humanity itself is implicit in every human choice. If complete normative skepticism is to be avoided—if there is such a thing as a reason for action—then humanity, as the source of all reasons and values, must be valued for its own sake. (1996c, 122)

I begin with three comments for clarification. First, although Korsgaard attributes this argument to Kant, that claim is controversial.² Some scholars (for example, Guyer 1998; Wood 1998; Langton 2007) interpret Kant differently. My aim in this article is to assess the two arguments that Korsgaard presents for the value of humanity (that is, what I am calling Kant's Argument and Korsgaard's Argument), so I set aside the exegetical question of whether Kant's Argument really is, as Korsgaard claims, *Kant's* argument. Thus, I will speak of Kant's Argument as being made by Korsgaard rather than by Kant.

Second, although Korsgaard refers in the quoted passage to “our desires and inclinations” as the explanation for why we find the objects of our choices good, she elsewhere emphasizes that this explanation is incomplete because we sometimes disapprove of our desires and their objects (1996a, 121). Strictly speaking, her claim is that we do, or should, regard the objects of

² Korsgaard (1996a) defends Kant's Argument at length as an interpretation of Kant's argument for the Formula of Humanity.

our choices as good because we choose them: “we are supposing that rational choice itself *makes* its object good” (1996a, 122). Desires are relevant because they present their objects as candidates for choice. If we endorse a desire and choose its object, then that object is good because we choose it. If we repudiate a desire, refusing to choose as it beckons, then its object is not good despite our desire for it. Korsgaard (1998, 51-52) holds that we typically choose to do what we desire to do unless another consideration, such as morality or prudence, dictates otherwise. So our desires can explain why something is good in a derivative way, by influencing our choices, which are “value-conferring” (1996a, 122).

Third, in both Kant’s Argument and Korsgaard’s Argument, Korsgaard assumes a connection between what people value and what really is valuable. The connection is roughly that if our valuing of some object is able to “withstand reflective scrutiny,” then that object is valuable (1996c, 93). To justify the connection, she relies on a metaethical claim about the nature of value: “Value . . . is only directly accessible from within the standpoint of reflective consciousness,” not from a third-personal perspective (1996c, 124). In other words, there is nothing more to humanity (or anything else) being valuable than our valuing it “in the full light of reflection,” just as there is nothing more to something being red than our seeing it as red under appropriate circumstances (1996c, 266). “Trying to actually see the value of humanity from the third-person perspective is like trying to see the colours someone sees by cracking open his skull. From outside, all we can say is why he sees them” (1996c, 124).

Though controversial, this metaethical claim is not unique to Kantian approaches. For example, Gibbard (1999, 153-54) endorses it as compatible with his expressivist metaethics. Moreover, Korsgaard’s arguments retain interest without it. A weaker *evidential* connection between valuing and value is plausible. Korsgaard argues that, at least after sufficient reflection,

everyone would come to value humanity as the source of all value. It is difficult to see how philosophical ethics could proceed without the assumption that the human capacity for evaluative reflection is at least roughly reliable. Given that assumption, convergence among all reflective agents in this respect is perhaps the best evidence we could have that humanity is valuable, however we understand value. Indeed, Korsgaard's conclusion, if true, would be important even if it provided no evidence at all about what really is valuable. If one holds humanity to be valuable on other grounds, then a psychological disposition to value humanity upon reflection justifies some optimism about moral progress by revealing a respect in which human nature is oriented toward the good. If one holds that humanity is not valuable, or valuable but not the source of all value, then the same psychological disposition would (also) be noteworthy as a bias or distorting influence on our reasoning—a sort of collective egocentrism to which reflective agents are constitutively prone. So I will grant Korsgaard's metaethical claim for the sake of argument.

With these clarifications in hand, we can reconstruct Kant's Argument a bit more formally as follows. I have named the premises for convenient reference:

1. We regard the objects of our choices as good. (Choice Premise)
2. If we regard the objects of our choices as good, then we are committed to regarding them as good because we choose them (rather than as good in themselves). (Subjectivism Premise)
3. If we regard something as good because we choose it, then we are committed to regarding ourselves as valuable. (Source Premise)

4. If we regard ourselves as valuable, then we are committed to regarding humanity as such (that is, others as well as ourselves) as valuable. (Generalization Premise)
5. If we are committed to regarding humanity as such as valuable, then humanity as such is valuable.³ (Metaethical Premise)
6. Therefore, humanity as such is valuable.

Critics have almost uniformly found Kant's Argument unpersuasive.⁴ "An unsympathetic reader may be tempted to view it as a chain of non sequiturs," says Langton (2007, 169), and even some very sympathetic readers (for example, Gibbard 1999) have found problems. Let's begin by setting aside three premises.

I grant the Metaethical Premise for the reasons already noted. What about the Choice Premise? It is plausible that, by choosing one option rather than another, I ordinarily at least implicitly endorse that option as better than the alternatives.⁵ Human choice is *rational* choice; we act for reasons. What about akratic actions, such as choosing to sleep in when I believe that it would be better to work instead?⁶ A hardline response would be to reject the possibility of akrasia. If I choose to stay in bed, then that is ipso facto what I value more, even if I tell myself otherwise. However, Korsgaard elsewhere rejects "the strange idea that an acknowledged reason can never fail to motivate" (1996b, 331). "Rage, passion, depression, distraction, grief, [and]

³ Earlier I described Korsgaard's metaethical claim as explaining value in terms of valuing that survives reflective scrutiny. Here I assume that if we are committed to valuing humanity, then our doing so would survive reflective scrutiny.

⁴ To repeat, by "Kant's Argument" I mean the argument that Korsgaard attributes to Kant, regardless of whether that argument accurately represents Kant's own philosophical intentions. Thus, if one takes Korsgaard to have misinterpreted Kant, then one can accept what one takes to be Kant's argument for the value of humanity while rejecting Kant's Argument.

⁵ For example, Gibbard (1999, 149): "Reflective endorsement is, in a sense, inescapable; if you think about what to do and come to a conclusion, you have thereby reflectively endorsed your conclusion. . . . And if I reflectively endorse, for instance, my impulse to have fun by swimming, I thereby *regard* its object (fun) as of value."

⁶ I thank an anonymous referee for encouraging more attention to this point.

physical or mental illness” can all disrupt the normal motivational connection between acknowledging a reason and acting accordingly (1996b, 320). Only if we are fully rational are we always moved appropriately by reasons. Korsgaard (1996a) does not explicitly refer to akrasia in working out Kant’s Argument, but her statements of the Choice Premise are similarly qualified: “Insofar as we are rational agents we will choose what is good—or take what we choose to be chosen as good” (1996a, 115). Thus, it seems best to interpret the Choice Premise as saying that, when we choose rationally in the sense that we are not overcome by interfering factors of the sort just mentioned (such as depression), we see the objects of our choices as good. This formulation leaves room for akratic action as a result of those interfering factors. (I assume nonakratic choice in what follows.)

I will also grant the Generalization Premise. That premise is crucial to Korsgaard’s defense of morality because it establishes the link between valuing one’s own humanity and valuing humanity as such. Without it, one could value one’s own humanity while remaining indifferent or hostile to others. In “Kant’s Formula of Humanity,” Korsgaard (1996a, 123) appears to defend the Generalization Premise on consistency grounds: if you value your own humanity, then you are committed to valuing the humanity of others as well, for there is no relevant difference between you and them. In *The Sources of Normativity*, Korsgaard (1996c, 132-34) rejects this strategy on the grounds that consistency is compatible with universal egoism, in which each person values only her own humanity and recognizes the legitimacy of others doing the same.⁷ Instead, she argues that reasons are “public in their very essence,” in the sense that other people’s reasons have normative force for us as if they were by default, rather than as the

⁷ Korsgaard (1996c, 134) still claims that consistency rules out regarding one’s own humanity as valuable without recognizing the legitimacy of others’ valuing their humanity in the same way. Thus, it would be inconsistent to expect everyone else to value my humanity while I do not value theirs.

conclusion of an argument (1996c, 135). To support that claim, she appeals to interpersonal phenomena such as our tendency to do what others request or else offer them a reason why we refuse, as well as the possibility of shared deliberation in which people reason together and reach a joint conclusion about what to do (1996c, 135-45). Korsgaard's argument on this point has been amply discussed elsewhere, so I will grant the Generalization Premise for the sake of argument in order to explore other issues.⁸ Even if Korsgaard could not establish that each person is committed to valuing humanity in others, to show that each person is committed to valuing her own humanity as the source of the value of everything she values would still be a significant accomplishment.

Now consider the Source Premise, which says that if we regard something as good because we choose it, then we are committed to regarding ourselves as valuable. Critics of Korsgaard often regard this inference as fallacious. It seems to be an application of a more general property transmission principle: if X confers property P on Y, then X must also have P. This principle has numerous counterexamples. For example, university presidents can confer doctoral degrees on graduating students even if they do not have such degrees themselves (Gaut 1997, 174).⁹ However, Korsgaard could plausibly argue that the Source Premise does not rely on the property transmission principle (1998, 64). She is already committed to rejecting that principle because she claims that choice can confer badness as well. For example, pain is "nearly always bad . . . because the creatures who suffer from it object to it" (1996c, 154). People can confer badness on something by rejecting it. But we should not conclude that people are bad!¹⁰

⁸ See, for example, Lebar 2001 and Gert 2002.

⁹ For similar counterexamples, see, for example, Langton 2007, 175-76.

¹⁰ Kerstein (2001, 34) presents the case of people conferring badness via reflective rejection as an objection to Korsgaard, but a more charitable interpretation is that it shows that she is not relying on the property transmission principle in the first place.

Instead of the property transmission principle, Korsgaard appears to rely on the idea that, if X matters, then what matters to X also matters. Thus, if the objects of your choices matter, then one candidate explanation for that fact is that *you* matter: “Kant urges us to take things to be important *because they are important to us*” (1996c, 242). Things can matter to someone in either a positive or a negative way, so if people matter, then they can confer both goodness and badness on the world.¹¹

Might the value we place on the objects of our choices be justified in some other way? I interpret the claim (in the Subjectivism Premise and the antecedent of the Source Premise) that we are committed to regarding the objects of our choices as good “because we choose them” as implying not only that we should not regard them as good in themselves, but also that we should not regard them as deriving their value from any third-party source that is sensitive to our choices. For example, suppose that God’s valuing confers value on its objects and that God values the objects of human choices (for example, out of benevolent concern). On this account, the fact that I choose something is part of the explanation of why it is good, but the account does not support the claim that I am valuable; instead, it supports the claim that God is valuable. If God does not value human beings, then human beings are not valuable, even if their choices are in this derivative way value-conferring. If alternative explanations of this sort were available, then the Source Premise might not be justified. However, as noted, I interpret the Subjectivism Premise as ruling out the claim that the objects of our choices derive their value from third-party sources that are sensitive to our choices. Thus, the claim that our choices are value-conferring

¹¹ The possibility of reflective rejection also solves a problem about interpreting the Choice Premise to allow for akrasia. The problem is: If someone always acts akratically, so that she never regards anything that she chooses as good, does she avoid the commitments that Korsgaard claims follow from doing so? But akratic action plainly involves regarding one’s choice as bad (even if also tempting in various ways). If the permanent akratic sees her reflective rejection of an option as capable of conferring badness on it, Korsgaard could argue, then she must take herself to be valuable.

because we ourselves are valuable seems to be the most plausible remaining explanation, in which case the Source Premise is justified.

The weight of Kant's Argument thus rests squarely on the Subjectivism Premise. It says that if we regard the objects of our choices as good, then we are committed to regarding them as good because we choose them, rather than as good in themselves. This premise has attracted the most ire from Korsgaard's critics. Many (for example, Gaut 1997, 182-83; Cohon 2000, 77-78; Regan 2002, 274-75) have argued that the Subjectivism Premise is at odds with ordinary evaluative thought and discourse. We can distinguish two objections. First, many people believe in standards of value that are independent of our desires and choices, and to which desire and choice should be responsive. On this view, some objects and activities are intrinsically better than others. Kant's Argument leaves room for objective moral standards, which derive from the value of humanity and properly constrain desire and choice. However, ordinary evaluative thought and discourse seems committed to nonmoral standards of value as well. For example, pushing a block of wood around the floor at precise intervals (Gaut 1997, 178) or counting blades of grass (Rawls 1999, 380) do not become valuable activities merely because one desires or chooses to engage in them. Call this the *objective standards objection*. Second, the conception of choice as value-conferring may be at odds with our experience of agency and choice (Regan 2002, 273-74; see also Enoch 2007, 35-41). If our choices determine what is good, then we have no criteria by which to choose (in nonmoral cases), so our choices must ultimately be arbitrary. But that seems false to our experience of choice, which frequently involves trying to make the *right* choice. Call this the *arbitrariness objection*.

Korsgaard has a plausible response to the arbitrariness objection. It relies on expanding the scope of the reasons that derive from the value of humanity to cover many choices that we

would ordinarily regard as nonmoral. First, Korsgaard (1998, 54) argues that if we value our humanity, then we should value satisfying our desires as an expression of “a sort of benevolence toward the self,” at least when doing so is not at odds with the value of humanity in other ways.¹² Thus, although we regard our choices as conferring value on their objects, our desires also provide choice-independent standards for action that properly guide choice on most occasions.¹³ Second, Korsgaard (2009, 209) argues that we typically operate within “traditions of value” created by the choices of others, so we can make choices at least partly on the basis of value already conferred by others. These considerations show that the Subjectivism Premise does not make our choices arbitrary in the context of Korsgaard’s theory, in which the value of humanity provides pervasive choice-independent standards for choice. However, the objective standards objection still applies, for the choice-independent standards just mentioned derive from our desires and from the choices of other people, not from the value of the objects and activities themselves. As Korsgaard (2009, 123) points out, this does not make their features irrelevant: the Kantian “can still talk to herself, and to others, about what she likes about them, and why.” But if someone wants to spend her time counting blades of grass, and does not neglect her moral duties

¹² Regan (2002, 279) argues that the defender of Kant’s Argument cannot appeal to desires to guide choice because “desire itself is essentially arbitrary, from the point of view of the free will.” As Korsgaard’s appeal to self-benevolence suggests, that claim is partly true and partly false. It is true that, on a Kantian view, a free will is guided first of all by the moral law, not by desire. But it is false that our desires have no implications for what we should do, for they may be morally relevant.

¹³ It may seem that Kant’s Argument does not even require the claim that we regard our choices as conferring value on their objects. The crucial point, one might think, is that we regard other things as valuable because they matter to us. What matters to us might be explained in terms of what we choose or in terms of what we desire, but either way, we can combine that claim with the Source Premise to get the conclusion that we must regard ourselves as valuable. However, I take it that Korsgaard is committed to the claim that choice in particular is value-conferring because she thinks that this affects the way in which we must value ourselves. In Korsgaard’s Kantian view, our humanity is our capacity for rational choice (that is, free choice, not determined by instinct), so she wants to show that we must value ourselves specifically qua rational choosers and then draw moral conclusions from this fact (1996a, 124-28). The claim that we regard the objects of our choices as good because we choose them supports that conclusion. If we regard our capacity for rational choice as value-conferring, then (by the Source Premise) we are committed to regarding that capacity—that is, our humanity—as valuable. In any case, a version of Kant’s Argument according to which we are committed to regarding our desires (rather than our choices) as valuing-conferring would also be vulnerable to the objective standards objection.

in order to do so, then Korsgaard is committed to the claim that she is not making a mistake when she values grass-counting.

In defense of the Subjectivism Premise, she writes:

Suppose that you make a choice, and you believe what you have opted for is a good thing. How can you justify it or account for its goodness? In an ordinary case it will be something for which you have an inclination, something that you like or want. Yet it looks as if the things you want, if they are good at all, are good because you want them—rather than your wanting them because they are good. For “all objects of inclinations have only a conditional worth, for if the inclinations and the needs founded on them did not exist, their objects would be without worth” (G 428). The objects of inclination are in themselves neutral: we are not attracted to them by their goodness; rather their goodness consists in their being the objects of human inclinations. (1996a, 121)

The argument seems to be:

1. We regard the objects of our choices as good. (Choice Premise)
 2. If something is an object of choice, then we have an inclination for it.
 3. If we did not have an inclination for it, then it would not be good.
 4. Therefore, if we regard the objects of our choices as good, then we are committed to regarding them as good because we choose them (rather than as good in themselves).
- (Subjectivism Premise)

But premise (3) is question-begging. If we lack an inclination for something, then granting premise (2), we would not choose it. But it follows that the would-be object of choice is not good only if we already presuppose the Subjectivism Premise, for the salient alternative is that the object might be good in itself, independently of our inclinations and choices. One might observe that if we lack an inclination for something, then we will not *regard* it as good, so what we regard as good and what we are inclined to do are coextensive. But that does not commit us to thinking that our inclinations and choices confer value on their objects. The order of explanation could go the other way: our inclinations and choices could track what we regard as good. One might observe that many things that we like—ice cream, hiking, the attendance of a butterfly upon a steamroller—are good only because we like them. There is no magic in ice cream. But as Korsgaard’s critics point out, there are other things, such as knowledge and art, which arguably do not depend for their value on our choices; and yet other things, such as counting blades of grass, that arguably are not good even if chosen (as we observe when other people, or our own past selves, choose them). Of course, it could be true on the basis of some other argument that the value of the former does depend on our choices and that the latter are indeed good if chosen. But as it stands, Kant’s Argument appears at best incomplete.

Is there an alternative? One might interpret Korsgaard’s argument differently, by treating the claim that the objects of choice are good because we choose them as a descriptive claim about how we regard our choices rather than a philosophical conclusion to which we are committed by the Choice Premise. In other words, Korsgaard could argue that we all at least implicitly regard our own choices as conferring value on an antecedently valueless world, and that attitude in conjunction with the Source Premise commits us to valuing ourselves (if we do

not already). Some of Korsgaard's language suggests this interpretation. Paraphrasing Kant, she claims that "in our private rational choices and in general in our actions we view ourselves as having a value-conferring status in virtue of our rational nature. We act as if our own choice were the sufficient condition of the goodness of its object: this attitude is built into (a subjective principle of) rational action" (1996a, 123). In a later summary of the argument, she says that "according to Kant the reason we treat our inclinations as reasons is not that we think the objects of inclinations are good in themselves. . . . It is rather that we take the objects of inclination to be good for us and we tend to take ourselves as good" (1998, 53-54). It sounds as if regarding our choices as valuing-conferring is not (or not only) something to which we are committed by the Choice Premise, but already a part of our pretheoretical experience.

The problem is that the philosophical critics cited above, who do not regard their choices in this way, are real-life counterexamples. *They* are not committed, on the basis of this version of Kant's Argument, to valuing humanity as the source of all value. One might respond by putting more weight on Korsgaard's claim that "any *reflective* agent can be led to acknowledge that she has moral obligations" (1996c, 125; my emphasis). On this construal, the problem with the philosophical critics is that they are insufficiently reflective. If they reflected more or better, then they would regard their choices as value-conferring. But this option leads to a dilemma. If better reflection would lead someone to accept this form of subjectivism because reflection leads people to believe the truth and this form of subjectivism is true, then we are back to the original interpretation of Kant's Argument and the problem that the argument for the Subjectivism Premise is unsound, or at best incomplete. Without some such argument in hand, however, we have no reason to believe that better reflection would indeed lead someone to regard her choices as value-conferring. The critics exhibit all the ordinary signs of sincere and thoroughgoing

reflection, so it would be ad hoc to claim that they are less reflective than the defenders of Kant's Argument, who do regard their choices in that way. Thus, on this construal, the Subjectivism Premise appears false: some reflective agents do not regard their choices in the way it describes.

Because the Subjectivism Premise receives no adequate defense, we may conclude that Kant's Argument for the value of humanity is unsuccessful. In the next section, we will see how Korsgaard's "fancy new model" of Kant's Argument has prima facie better prospects because it supplies a novel argument for the Subjectivism Premise. It retains the idea, just suggested, that sufficiently reflective agents will not value other things for their own sake but instead regard their doing so as justified by their own desires and choices (or those of others) and ultimately by the value they place on humanity. However, in Korsgaard's Argument, this idea does not follow simply from the Choice Premise or from observation of actual people but instead from a broader theoretical account of rational agency.

2. Korsgaard's Argument

I will present Korsgaard's Argument in two parts. The first part is a conception of rational agency with three important features. The second part is an argument that any rational agent so understood will, after sufficient reflection, value humanity as the source of all value. I will also explain the main advantage of Korsgaard's Argument over Kant's Argument. In the next section, I will show how the argument is nevertheless unsuccessful.

Korsgaard's Argument relies on a conception of rational agency with three important features. The first is the *reflective structure of human consciousness*. Unlike other animals, Korsgaard argues, humans are conscious of their own mental states and, as a result, face the question of whether to endorse them as a basis for action or belief (1996c, 92-93). We have the

capacity to step back from our desires and reflectively evaluate them. Moreover, the reflective structure of human consciousness does not merely give us the option of endorsing or rejecting desires that, if we do not intervene, will produce action on their own. Instead, our mind is such that we must endorse a desire before we can act at all. The life of Frankfurt's (1971, 11-14) wanton, who is moved by whatever first-order desires he happens to have, without any second-order attitude of endorsement, is not possible for us (Korsgaard 1996c, 99 n. 8). As a result, we face "a problem no other animal has" (1996c, 93). We must determine which of our desires to endorse before we can act. We will see Korsgaard's solution to the problem shortly.

The second important feature of rational agency is the *necessity of action* (2009, 1-2). Korsgaard claims, plausibly enough, that we do not have the option of directly "opting out" of agency and action. We can temporarily or permanently eliminate the need for action in the future by taking certain actions now, such as swallowing sleeping pills or committing suicide. We can also lose our capacity for action in other ways, such as if we fall asleep or someone knocks us unconscious. But here and now, awake and functioning normally, we have no option but to act in one way or another, even if what we choose to do is nothing. The necessity of action is important because it makes the problem posed by the reflective structure of human consciousness more serious. It means that we cannot avoid the problem of determining which desire to endorse by declining to act at all. If we literally must act, and we must endorse a desire in order to act, then we must endorse a desire. "The necessity of choosing and acting," says Korsgaard, "is our *plight*: the simple inexorable fact of the human condition" (2009, 2). Because of the aforementioned possibilities of sleep, suicide, and the like, the plight is neither constant nor inescapable, but its bounds define the bounds of active human life.¹⁴

¹⁴ An anonymous referee suggests that "a depressed or apathetic person may encounter the opposite plight, of utter inertia and apathy, with no necessity for choosing or acting." I take it that most cases of depression are not like this

The third important feature of rational agency is the *need for justification*. Korsgaard argues that we cannot solve the problem posed by the reflective structure of human consciousness by arbitrarily endorsing one desire rather than another. Instead, we must have a justification for endorsing a particular desire. This is clear from how Korsgaard describes the problem: “The reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*. Otherwise, at least as long as it reflects, it cannot commit itself or go forward. . . . We need reasons because our impulses must be able to withstand reflective scrutiny. We have reasons if they do. The normative word ‘reason’ refers to a kind of reflective success” (1996c, 93). And, “‘Reason’ means reflective success. So if I decide that my desire is a reason to act, I must decide that on reflection I endorse that desire. And here we run into the problem. For how do I decide that?” (1996c, 97).

Korsgaard does not explain, at least in these passages, why we cannot endorse desires arbitrarily, without regarding them as justified. However, her reference to “withstanding reflective scrutiny” suggests an explanation in terms of the reflective structure of human consciousness. We have the power, and the liability, to step back from our desires and ask whether they provide reasons for action. An arbitrary endorsement of a desire would not meet the demands of reflection because it would itself be just another psychological impulse. As reflective, we would immediately step back from our arbitrary endorsement and ask whether to

because, even if depression leaves one psychologically unable to do many of the things that one wants to do, it does not wholly disable choice and action. Perhaps I cannot bring myself to go to work or meet my friends, but I can (and must) still choose which meal to eat or television program to watch, or (minimally) what to think about next. A depression so severe that it left one literally unable to choose or act at all, even in the most minimal ways, would be—as Korsgaard’s claim suggests—a kind of falling-away from human life. There is indeed an opposite plight here: the capacity to choose and to act is surely a power as well as a burden, and part of the viciousness of even less severe forms of depression is to find oneself unable to choose and act in ways that, by all external appearances, are available to one.

endorse *it*. If so, then arbitrary endorsement could not satisfy a reflective consciousness. To endorse a particular desire as a basis for action, we must regard our endorsement as justified.

What criteria of justification do we apply in reflection? Korsgaard argues that one's *practical identity* provides the criteria. A practical identity is "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking" (1996c, 101). It can be understood as a set of principles or more abstractly as "a role with a point" (2009, 21). For example, the practical identity of a citizen includes the principle of paying one's taxes because it is the law that one do so (1996c, 106). These principles and points provide criteria by which we can determine whether to endorse a given desire: "We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves" (1996c, 120). If I identify as a citizen, then I will endorse my desire to pay my taxes, and repudiate my desire not to. We do not always think explicitly in terms of our practical identities when we reflect, though we sometimes do, especially in the face of temptation to act otherwise (2009, 21). Korsgaard concludes: "Your reasons express your identity, your nature; your obligations spring from what that identity forbids" (1996, 101).

To summarize: Korsgaard's conception of rational agency includes three important elements: (1) the reflective structure of human consciousness, which sets us the problem of deciding which desires to endorse as a basis for action; (2) the necessity of action, which makes that problem unavoidable; and (3) the need for justification, which constrains possible solutions by ruling out arbitrary endorsement in favor of the criteria provided by one's practical identity. Now, one might reject this conception of rational agency. However, each feature has intuitive plausibility, so it is worth granting Korsgaard's conception of rational agency in order to see

what follows from it. If it were true that agents of this sort were inevitably committed, upon reflection, to valuing humanity as the source of all value, then that would be an astonishing discovery even if we remained uncertain about whether we are such agents.

As Korsgaard recognizes, the story so far is compatible with a kind of relativism (1996c, 113). Her ultimate aim is to show that everyone is committed to valuing humanity as the source of all value. I will continue to grant the Generalization Premise: if we value our own humanity, then we must also value the humanity of others. However, so far, it seems that some people might not value themselves as human but only in terms of more particular practical identities such as their family, professional, national, or religious identities. For example, Korsgaard acknowledges the possibility of a mafioso who values himself under that description, with its associated principles of loyalty to insiders and exploitation of or brutal retribution against outsiders (1996c, 254-58). Moreover, even those who do value their humanity may not regard it as what justifies their other practical identities. They might value their own family or nation more than humanity as such. That would be a problem for Korsgaard because she wants to argue that the reasons and obligations deriving from humanity—that is, moral reasons and obligations—generally, though perhaps not always, trump those of competing practical identities (1996c, 125-26). The fact that the value we place on other practical identities derives from the value we place on humanity would explain why the latter takes priority in cases of conflict. So why does Korsgaard think that every sufficiently reflective agent will value humanity as the source of all value?

We have seen that we determine which desires we should endorse by appealing to the criteria provided by our practical identities. However, a crucial feature of the reflective structure of human consciousness, on Korsgaard's view, is that it applies in an iterative or regressive

fashion. We can justify one claim by appealing to another, but then we can step back from the second claim and ask what justifies *it*, and so on. Thus, just as we step back from our desires and ask what justifies them, we also step back from our practical identities and ask what justifies *them*. “Why should it matter whether I live up to the demands imposed upon me by citizenship, or motherhood, or my profession?” (1996c, 129). The iterative nature of reflection precludes justifying one’s particular practical identities by appealing to (1) their intrinsic features or (2) other practical identities (with one exception below). For example, Cohon (2000, 77-78) objects to Korsgaard that reflection can properly conclude with justifications that appeal to the intrinsic features of one’s practical identity. One might justify one’s practical identity as a musician by appealing to the value one places on music, say. But given the reflective structure of human consciousness, that will not work: a reflective agent must go on to ask whether valuing music is justified.

It is worth noticing that Korsgaard does not rely on the claim that justifications of this sort are inadequate on *epistemic* grounds. Critics (for example, Cohen 1996, 179-180; Gibbard 1999, 146) sometimes complain that Korsgaard erroneously assumes that an adequate answer to our evaluative questions must be capable of persuading a “radically disaffected” skeptic. On this interpretation, Korsgaard regards “substantive realist” accounts of morality, according to which moral principles are irreducibly normative, as inadequate because a moral skeptic would reject them. The critics respond that satisfactory justifications for morality and other values are available even if they would not persuade the skeptic. As Gibbard (1999, 164) puts it, the substantive moral realist “can look for deep, underlying principles that are plausible, and aim to show that no coherent alternative to morality remains plausible on examination.” Someone unmoved by justificatorily basic and intuitively plausible principles, such as the principle that

one's enjoyment of doing something counts in favor of doing it, may need "therapy" rather than philosophical arguments (Gibbard 1999, 146).

However, in my view, this interpretation fails to appreciate how the reflective regress arises from the iterative nature of the reflective structure of human consciousness in conjunction with the need for justification. The problem faced by the reflective agent is not that one or another justification for morality or another value happens to fall short according to some independent epistemic standard (such as coherence), but rather that she can apparently step back from *any* proposed justification and ask what justifies *it*. In this sense, her problem is agential, not epistemic. It is reflection itself that is radically disaffecting, at least if one cannot answer the questions it raises in such a way as to end the reflective regress. One might reject Korsgaard's conception of rational agency in favor of the view that, when we reflect, we accept some commitments as basic rather than stepping back from them and demanding that they too be justified. But we have granted Korsgaard's account of rational agency for the sake of argument, so the important point here is simply that the reflective regress that Korsgaard describes follows naturally from it.

What would end the reflective regress, Korsgaard (1996c, 33) argues, is an "unconditional" answer, "one that makes it impossible, unnecessary, or incoherent to ask why again." Only with such an answer could we fully solve the problem created by the conjunction of the reflective structure of human consciousness, which requires that we endorse desires before acting; the necessity of action, which makes action and hence endorsement unavoidable; and the need for justification, which precludes arbitrary or unjustified endorsement. Moreover, because (Korsgaard claims) these three features are constitutive of rational agency, any rational agent

faces this problem. If it has a unique solution, then that solution will have authority for every rational agent. It is what every agent would arrive at after sufficient reflection.

Korsgaard's proposed solution is that we value our humanity and regard it as justifying our other, more particular practical identities. Here is the crucial passage:

Most of the self-conceptions which govern us are contingent. . . . 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 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.

(1996c, 121)

We can reconstruct the argument a bit more formally as follows:

1. We must find some justification for valuing our particular practical identities.
2. Because we are human, we need reasons in order to act and to live.
3. If we value our humanity, then it can justify whatever satisfies our need for reasons.
4. Our particular practical identities satisfy our need for reasons.

5. Therefore, if we value our humanity, then it can justify our particular practical identities.
6. We have no other justification for valuing our particular practical identities.
7. Therefore, we must value our humanity.

This argument has two main parts. One part, represented by premise (6), is the earlier conclusion that when reflecting we cannot ultimately regard our particular practical identities as justified either by their intrinsic features or by other particular practical identities. Because of the iterative nature of the reflective structure of human consciousness, a reflecting agent will step back from any proposed justification and demand a further justification of it, apparently without limit. Nevertheless, we do value our particular practical identities, so given the need for justification, we must find some justification for them, as premise (1) says.

The second part of the argument is that one special practical identity, our identity as human, can provide that justification. Because we are human and thus characterized by the three features of rational agency, we need some practical identity or other in order “to act and to live” (1996c, 121). Recall: the necessity of action means that we must act; the reflective structure of human consciousness means that we must endorse desires in order to act; the need for justification means that we cannot endorse desires arbitrarily or without justification. We need reasons. If we value our humanity, then we can value whatever meets that need for reasons. Our particular practical identities meet this need by providing criteria for the reflective endorsement of desires, and thus reasons for action.¹⁵ Thus, because we do value our particular practical identities, and because we have no other justification for them, we will upon sufficient reflection

¹⁵ Why do we need other practical identities besides humanity? Korsgaard suggests that they will continue to govern many of our choices even after we accept humanity as a foundational value (1996c, 125). This is presumably because the value of humanity provides moral criteria for action, but morality does not dictate what to do on every occasion for choice, so we need other practical identities to fill the gaps.

come to value our humanity (if we do not already) and regard it as what justifies our other practical identities.¹⁶ There is elegance here: the nature of humanity (that is, the three features of rational agency) sets us a problem (that is, how we should act), and the value we place on humanity solves it (by justifying our particular practical identities, which provide us with reasons for action).¹⁷

Could one value one's particular practical identities on the grounds that they meet one's need for reasons, but not value one's humanity? If some other practical identity prescribes meeting that need, then perhaps it could explain why one values meeting it. For example, a hedonist might realize that she is most happy when she knows what to do and most unhappy when uncertain or bored. Satisfying her need for reasons is essential to her happiness. She values meeting that need on that basis and adopts some other practical identities that provide her with criteria for acting on all sorts of occasions. But she never values her humanity as such, and if her life no longer promised a balance of pleasure over pain, then she would willingly destroy her humanity. However, alternative justifications of this sort are unavailable in the present context because they succumb to the reflective regress described earlier. The hedonist would have to step

¹⁶ Strictly speaking, Korsgaard (1996c, 122) says that the value we place on other practical identities is "partly" derived from the value we place on humanity: "other forms of practical identity matter in part because humanity requires them." She does not explain why else they matter, but presumably she has in mind the personal and historical considerations that explain why we have some practical identities rather than others. Kerstein (2001, 47-48) objects that Korsgaard's theory does not provide a basis for choosing one practical identity rather than another. But recall Korsgaard's (1998, 54) claim that valuing our humanity requires giving weight to our own interests and desires as an expression of "a sort of benevolence towards the self," at least when doing so is not in other ways at odds with valuing humanity. Our desires thus provide a justification for choosing one practical identity over another even when either would satisfy our need for reasons.

¹⁷ FitzPatrick (2005, 678-81) argues that Korsgaard's claim that the value one places on other practical identities derives from the value one places (or is committed to placing) on humanity succumbs to a dilemma. If she means that everyone literally must regard their other practical identities as justified in this way, then her claim is obviously false, since many people clearly do not. However, if she means that valuing other practical identities commits one to valuing humanity because it is *true* that the value of other practical identities derives from the value of humanity, then she needs to provide independent argument for that claim, and she does not. However, FitzPatrick neglects the role of the reflective regress, which supports the commitment claim not by showing that it is true that the value of other practical identities derives from the value of humanity, but by showing that any sufficiently reflective agent would come to regard her practical identities in this way, because (Korsgaard argues) no other justification of them can end the reflective regress. Thus, Korsgaard avoids the dilemma by taking a third option.

back from her hedonism and ask what justifies it, and so on, until (if Korsgaard is right) she is driven to appeal to her need for reasons in order to act and to live. Because she cannot at this point explain the significance of that need by appealing to her hedonism (which is what she is now trying to justify), her only alternative is to value her humanity.

Why does Korsgaard's Argument have an advantage over Kant's Argument? Recall the Subjectivism Premise from Kant's Argument. It says that if we regard the objects of our choices as good, then we are committed to regarding them as good because we choose them (rather than as good in themselves). We saw that the latter claim is open to two objections: (1) the objective standards objection, which says that some objects of choice are valuable independently of our desires and choices, and (2) the arbitrariness objection, which says that choice requires choice-independent standards of value. I argued that Korsgaard can plausibly respond to the arbitrariness objection by arguing that her theory provides adequate choice-independent standards because, in addition to obviously moral reasons, we have moral reasons of self-benevolence to do what satisfies our desires unless doing so conflicts with the value of humanity in other ways, and because we can also respond to the value that other people confer on the world. However, she must reject the objective standards objection as misguided.

Now, all the same points apply in the context of Korsgaard's Argument. On her view, a sufficiently reflective agent will not value her practical identities (besides humanity) for their own sake. Instead, she will value them on the grounds that they satisfy her need for reasons. This need does not justify some practical identities over others. As long as they provide sufficiently extensive criteria for choice, any set of practical identities is as good as any other. However, once that argument leads us to value humanity, then self-benevolence and the value-conferring choices of others provide choice-independent standards for adopting particular practical

identities, so our choice of practical identities need not be arbitrary. But Korsgaard must reject the objective standards objection as misguided: besides humanity, we can regard no practical identities as justified in themselves, apart from desire and choice. More precisely, although our need for reasons does justify (independently of our desires and choices) having some practical identity rather than none, what justifies any particular practical identity over another is simply that we (or other people) desire or choose it. If we want to, we have as much reason to become a grass-counter as a scientist or an artist. The crucial difference is that Korsgaard's Argument contains a powerful argument for its version of the Subjectivism Premise. Instead of arguing that the conception of choice as value-conferring follows simply from the coextension of what we are inclined to do and what we regard as good, or alternatively that everyone already at least implicitly regards their choices in this way, Korsgaard argues that no other attitude (that is, regarding one's practical identities as justified by their intrinsic features) would survive the reflective regress.

What about the philosophical critics who were apparent counterexamples to the latter argument for the Subjectivism Premise—that is, apparently reflective people who do not regard their desires and choices as conferring value on their objects? We noted that it would be ad hoc for Korsgaard to deny that the critics are sufficiently reflective, for we had no independent grounds for thinking that they are less reflective than the proponents of Kant's Argument. However, Korsgaard's Argument provides just such independent grounds in the form of her account of reflection based on the three features of rational agency. That account allows us to explore the dynamics of reflection in a theoretical way not in hock to observation of actual people, whose reflection may be more or less incomplete or inadequate. If her account is correct, then we can say that sufficiently reflective agents—that is, those who reflect long enough and

well enough to uncover what Korsgaard claims is the only justification for one's practical identities capable of withstanding reflective scrutiny—would come to value humanity as the source of all value. If some actual people do not reach that conclusion, then it is not ad hoc to regard them as insufficiently reflective rather than as counterexamples.

3. Critical Assessment of Korsgaard's Argument

Korsgaard's Argument is a marked improvement over Kant's Argument. However, the argument fails on its own terms: it fails to show that a sufficiently reflective agent must value humanity as the source of all value. To show why, I will first present the final piece of Korsgaard's Argument and then explain the mistake. The final piece of the argument is motivated by the question: what justifies valuing humanity? Cannot the reflective agent question that practical identity like any other? Korsgaard suggests two responses. First, she writes:

The price of denying that humanity is of value is complete practical normative scepticism. . . . Is there an argument against that kind of scepticism, a reason not to commit suicide? In one sense, the right reply is that there can't be, since life itself is the source of reasons. In another sense, the right reply is that this is not an issue to be settled by philosophical argument alone . . . value only exists if life is worth living, and that depends on what we do. (1996c, 163)

At this point, Korsgaard assumes that valuing humanity ends the reflective regress if anything does, so the question is whether to value humanity or nothing. She says that this

question is not settled by philosophical argument alone. Instead, we will value humanity if we find life worth living, which depends on what our life is like.

This argument confronts a dilemma. Korsgaard may intend to argue that one's conviction that life is worth living can end the reflective regress. But that claim is incompatible with her premises. Given the reflective structure of human consciousness and the need for justification, a reflective agent must step back from her attitude of valuing life and ask whether it is justified. A commitment grounded in recognition of the good things in life, such as one's personal relationships and professional achievements, merely returns one to an earlier stage of the reflective regress: why should we value our particular practical identities? Moreover, if a brute commitment to valuing life were somehow beyond the need for justification, then the same would apparently be true of a brute commitment to valuing other practical identities, such as parenthood. In that case, valuing humanity would not uniquely end the reflective regress, so Korsgaard's argument would not show that any sufficiently reflective agent would value humanity, rather than some other practical identity, as the source of all value. On the other hand, if Korsgaard does not intend to argue that one's conviction that life is worth living can end the reflective regress, then the question of how valuing humanity does end the reflective regress remains unanswered.

However, Korsgaard also suggests a second and *prima facie* more plausible argument. She agrees that we can question whether to value our humanity, but she argues that in this case we can answer the "why?" question in a way that ends the reflective regress.¹⁸ She writes:

¹⁸ Note that trying to justify the value of humanity in terms of some other practical identity would not end the reflective regress because we could then ask why *that* identity is justified.

Now that you see that your need to have a normative conception of yourself comes from your human identity, you can query the importance of that identity. . . . Does it really matter whether we act as our humanity requires, whether we find some ways of identifying ourselves and stand by them? But in this case you have no option but to say yes. Since you are 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, then you could have no reasons for action, and because your consciousness is reflective, you could then not act at all. 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. (1996c, 123)

We can reconstruct the argument a bit more formally as follows:

1. We must act.¹⁹
2. We must value some practical identity in order to act.
3. Therefore, we must value some practical identity.
4. The value we place on any other practical identity derives from the value we place (or are committed to placing) on our humanity.
5. Therefore, we must value our humanity.

¹⁹ This premise is implicit in the quoted passage when Korsgaard argues that “you *must* take something to be normative” because otherwise “you could . . . not act at all” (1996c, 123). Not acting at all is not an option, in her view.

Premises (1) and (2) reflect the three features of rational agency. In order to act, we must have some criteria by which to endorse or reject desires: “You cannot act without reasons” (1996c, 123). Our practical identities provide these criteria. Because we literally must act, we literally must value some practical identity or other. Premise (4) invokes Korsgaard’s earlier claim that our need for reasons in order to act is what justifies our more particular practical identities. If we value our practical identities because they meet that need, then we are committed to valuing our humanity; otherwise our needs qua human would not justify anything. In this way, the value we place on other practical identities derives from the value we place (or are committed to placing) on humanity: “your humanity is the [ultimate] source of your reasons” (1996c, 123).

In short, Korsgaard’s argument here is that we must act, and we must value some practical identity or other in order to act, but the value we place on any other practical identity derives from the value we place (or are committed to placing) on humanity, so we must value humanity—so humanity is valuable.²⁰ In this way, valuing our humanity “brings a regress of justification to a satisfactory end” (1996c, 111). We can coherently question whether to value our other practical identities because we can give them up. But we cannot coherently question whether to value humanity because the only alternative is to give up on every practical identity and thus on action altogether, and that is literally impossible.

Strictly speaking, Korsgaard (1996c, 163) claims not to have shown that “*complete* normative skepticism is impossible.” If people “feel that they themselves are worthless and, as a result, that life has no meaning and nothing is of value,” they can commit suicide rather than valuing humanity (1996, 162). Langton (2007, 185) objects to this feature of Korsgaard’s view on the grounds that such people are mistaken to place no value on their lives: “The emptiness in

²⁰ The last point follows with the addition of the Generalization and Metaethical Premises, which I continue to grant here (as in Kant’s Argument) for the same reasons.

[their] valuing is not an emptiness in [their] value.” I think that, if it works, Korsgaard’s Argument supports Langton’s claim, as well as the Kantian claim that suicide is at least *prima facie* immoral because at odds with the value of humanity.²¹ What Korsgaard should say is that someone who regards herself as worthless is either insufficiently reflective or in some way irrational.²² Though one can of course coherently avoid valuing one’s humanity by being dead, one cannot (at least in the full light of reflection) coherently kill oneself for that reason, for—if Korsgaard’s Argument is sound—even that action would commit one to valuing one’s humanity and thus to rejecting suicide, at least in the absence of some morally respectable reason in its favor.²³

However, the argument is unsound. The problem is premise (4), according to which the value we place on any other practical identity derives from the value we place (or are committed to placing) on humanity. That premise is question-begging at this stage of Korsgaard’s

²¹ See Velleman 1999 for a recent Kantian argument against suicide.

²² Korsgaard also has other grounds for resisting Langton’s objection. Langton describes the real-life case of Kant’s contemporary Maria von Herbert, who came to believe after a disappointed love affair that life and she herself were without value, and subsequently committed suicide. Langton objects that if Korsgaard’s theory were true, then Maria would indeed lack value, since she does not value herself. Two replies. First, even if Maria does not value much, she does appear to value a few things: the lover she disappointed (the relationship she might have had, by comparison with which a friendship “doesn’t have any point”), the prospect of visiting Kant, the prospect of ending her life, and morality itself, which she is concerned forbids suicide in her case (Langton 1992, 193). On Korsgaard’s account, even one of these interests commits Maria to valuing herself. Only the total nihilist (a much rarer character, perhaps even a philosophical fiction) raises the alleged problem for Korsgaard’s theory. On the latter point, Langton (2007, 181 n. 35) suggests that only a philosopher could conceivably deny that it is possible to value nothing, but her thought may be motivated by a grander conception of valuing than the one Korsgaard employs, in which recognizing any reason for action, no matter how trivial the occasion (for example, I eat when I am hungry), launches the reflective regress. Second, even if Maria were a total nihilist, it does not exactly follow that she is without value. For even if she does not confer value on herself, she may still have value conferred on her by others. If Korsgaard is right, the rest of us are rationally committed to valuing humanity as such, including Maria. This value is not something to which Maria herself must be responsive, but to the extent that Langton’s objection gets its power from the thought that, on Korsgaard’s view, when Maria proposes suicide, the rest of us should simply shrug and smile (“we should not shed tears”), it is mistaken (Langton 2007, 181). Relative to us, it is true that she is valuable and that her death was the loss of something priceless. That said, Langton’s larger objection that on Korsgaard’s view people have conditional rather than unconditional value—conditional, that is, on their valuing something—is, I take it, accurate. The question is whether that result is a problem if the condition is less easily avoided and, when unmet, less pregnant with implication for other people than Langton supposes.

²³ Korsgaard (1996c, 161–62) allows that suicide is sometimes justified when it is called for by one’s practical identity. This kind of suicide is a way to protect, promote, or honor what one values, whereas the kind of suicide in question in the text is a way to avoid valuing humanity or anything else.

Argument. She is arguing that a sufficiently reflective agent must value humanity. She first argued that valuing humanity allows one to justify one's other practical identities (except for those essentially in conflict with it). As human, we need some practical identity or other; if we value our humanity, then this need matters, and we can justify other practical identities on the grounds that they satisfy it. But then she encountered a problem: we can apparently question whether we should value our humanity. Until we have a satisfactory answer to that question, we cannot assume that valuing humanity is justified. And—this is the crucial point—if we cannot assume that valuing humanity is justified, then we cannot assume that the value we place on other practical identities derives from the value we place (or are committed to placing) on humanity.

Why not? The only reason for thinking it did was that valuing humanity seemed to offer a better justification for our other practical identities than justifications that appealed to their intrinsic features or to other practical identities besides humanity. To make good on that claim, Korsgaard needs to show that valuing humanity can end the reflective regress in a way that other potential justifications cannot. It would be circular to assume in the course of doing so that the value we place on other practical identities derives from the value we place (or are committed to placing) on humanity. If that claim is true, then it is true *because* valuing humanity can end the reflective regress, so it cannot without circularity be used in an argument to *show* that valuing humanity can end the reflective regress. Thus, premise (4) of Korsgaard's argument for that claim begs the question.

Of course, it may be true that, because we must act, we must value some practical identity or other. That was premise (3). But because we cannot assume that the value we place on other practical identities derives from the value we place (or are committed to placing) on

humanity, it does not follow that we are committed to valuing humanity in particular. We could coherently value some other practical identity without regarding it as justified by any value we might place on our humanity. Thus, Korsgaard has not shown that valuing humanity can end the reflective regress. Like any other practical identity, we can question whether it is justified and reject it if no adequate justification is forthcoming. A reflecting agent, if she is to act, must either find some other solution to the reflective regress or, if there is none, endorse some practical identity or other as sufficiently well-justified despite the possibility and coherence of further questioning.

Can we repair the argument? Recall Korsgaard's earlier claim that, if we assume that something is justified because it satisfies a need of ours, then we implicitly value ourselves. Korsgaard might argue that, if we treat the necessity of action and the need for justification as grounds for valuing something or other, then we are treating what we need as something that matters, so we are implicitly valuing our humanity, because it is by virtue of our humanity that we need reasons in order to act. However, Korsgaard's observation does not apply here. Although the necessity of action and the need for justification have the result that we literally must value something or other, they do not entail that we must value it *on any particular basis*, including on the basis that we are the kind of being that needs reasons in order to act. The need for justification will push us, in reflection, to find a justification for whatever particular practical identity we come to adopt. But such reflection will not necessarily lead us to value our humanity if valuing our humanity would not in fact uniquely justify that other practical identity. Korsgaard aims to show that it does because it alone can end the reflective regress, but as we have seen, her argument for that claim begs the question by assuming that the value we place on other practical identities derives from the value we place on humanity.

It is worth noticing that Korsgaard's other suggestion (discussed at the beginning of this section) about how valuing humanity can end the reflective regress is question-begging in the same way. She (1996c, 163) argued that "the price of denying that humanity is of value is complete practical normative scepticism" and that at least most people reject normative skepticism because they find life worth living, so they will favor the option of valuing humanity. I argued that this explanation fails because a brute commitment to valuing life cannot end the reflective regress on Korsgaard's own terms; a reflective agent must ask whether valuing life is justified. But whether we face that choice—either valuing humanity or valuing nothing—in the first place depends on Korsgaard's claim that all other possible values derive their value from the value of humanity. But if that claim is true, then it is true *because* valuing humanity can end the reflective regress, so it cannot without circularity be used in an argument to *show* that valuing humanity can end the reflective regress. So even if it had no other problems, Korsgaard's appeal to our conviction that life is worth living would fare no better than her appeal to the necessity of action.

The upshot is that Korsgaard's Argument suffers from a purely internal problem—a logical fallacy—that even her most sympathetic readers should acknowledge. It begs the question in favor of the value of humanity and its putative role in justifying other practical identities. It does so even if we grant Korsgaard various controversial premises, including her account of value in terms of valuing that survives reflective scrutiny (that is, the Metaethical Premise), her claim that valuing one's own humanity commits one to valuing the humanity of others (that is, the Generalization Premise), and her conception of rational agency, including the iterative nature of the reflective structure of human consciousness and the need for justification.

Kantians who wish to continue the project of showing that rational reflection commits everyone to valuing humanity as the source of all value must show how to repair Korsgaard's arguments or else look elsewhere for support.

References

- Cohen, G. A. 1996. "Reason, Humanity, and the Moral Law." In *The Sources of Normativity*, ed. Onora O'Neill, 167-188. New York: Cambridge University Press.
- Cohon, Rachel. 2000. "The Roots of Reasons." *Philosophical Review* 109: 63-85.
- Enoch, David. 2007. "An Outline of an Argument for Robust Metanormative Realism." In *Oxford Studies in Metaethics*, vol. 2, ed. Russ Shafer-Landau, 21-50. New York: Oxford University Press.
- FitzPatrick, William J. 2005. "The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity." *Ethics* 115: 651-691.
- Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68: 5-20.
- Gaut, Berys. 1997. "The Structure of Practical Reason." In *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut, 161-188. New York: Oxford University Press.
- Gert, Joshua. 2002. "Korsgaard's Private-Reasons Argument." *Philosophy and Phenomenological Research* 64: 303-324.
- Gibbard, Allan. 1999. "Morality as Consistency in Living: Korsgaard's Kantian Lectures." *Ethics* 110: 140-164.
- Gowans, Christopher W. 2002. "Practical Identities and Autonomy: Korsgaard's Reformation of Kant's Moral Philosophy." *Philosophy and Phenomenological Research* 64: 546-570.
- Guyer, Paul. 1998. "The Value of Reason and the Value of Freedom." *Ethics* 109: 22-35.
- Kerstein, Samuel J. 2001. "Korsgaard's Kantian Arguments for the Value of Humanity." *Canadian Journal of Philosophy* 31: 23-52.

- Korsgaard, Christine M. 1996a. "Kant's Formula of Humanity." In *Creating the Kingdom of Ends*, 106-132. New York: Cambridge University Press.
- 1996b. "Skepticism about Practical Reason." In *Creating the Kingdom of Ends*, 311-334. New York: Cambridge University Press.
- 1996c. *The Sources of Normativity*. New York: Cambridge University Press.
- 1998. "Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind." *Ethics* 109: 49-66.
- 2009. *Self-Constitution: Agency, Identity, and Integrity*. New York: Oxford University Press.
- Langton, Rae. 1992. "Duty and Desolation." *Philosophy* 67: 481-505.
- 2007. "Objective and Unconditioned Value." *Philosophical Review* 116: 157-185.
- LeBar, Mark. 2001. "Korsgaard, Wittgenstein, and the Mafioso." *Southern Journal of Philosophy* 39: 261-271.
- Rawls, John. 1999. *A Theory of Justice*. Rev. ed. Cambridge, MA: Belknap Press.
- Regan, Donald H. 2002. "The Value of Rational Nature." *Ethics* 112: 267-291.
- Street, Sharon. 2012. "Coming to Terms with Contingency: Humean Constructivism about Practical Reason." In *Constructivism in Practical Philosophy*, ed. James Lenman and Yonatan Shemmer, 40-59. New York: Oxford University Press.
- Velleman, J. David. 1999. "A Right of Self-Termination?" *Ethics* 109: 606-628.
- Wood, Allen W. 1998. "Review." *Philosophical Review* 107: 607-611.