



Review

Author(s): Sergio Tenenbaum

Review by: Sergio Tenenbaum

Source: *Ethics*, Vol. 121, No. 2 (January 2011), pp. 449-455

Published by: [University of Chicago Press](#)

Stable URL: <http://www.jstor.org/stable/10.1086/657967>

Accessed: 22-12-2015 00:51 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Ethics*.

<http://www.jstor.org>

point is that it would provide explanations of personal reasons in non-personal terms without going so far as to identify reason and morality in the way Hurley does.

Indeed, on this other sort of view, the rational authority of moral standards would ultimately depend on people's personal reasons for exhibiting a robust form of equal concern for one another. It seems quite unlikely, therefore, that people would typically have decisive reasons to act morally. However, it could still turn out that they would often have significant reasons to act morally and, thus, that a weaker claim about the rational authority of moral standards could be vindicated. For it might be easy for people to develop strong personal reasons of the requisite sort.

Be this as it may, my worry is that Hurley, having marginalized agent-neutral reasons, makes it all too easy to argue that personal reasons must be explained in terms of the sorts of impartial and interpersonal reasons with which he has—I think rightly—identified morality. Perhaps it's true that agent-neutral reasons can't be basic, but that needs to be established—and, especially in the context of a book attempting to persuade us to move beyond consequentialism, it needs to be established through arguments more charitable to consequentialism than the ones Hurley offers.

In closing, though, I would like to say again that this really is an impressive book. It provides the most rigorous argument that I have seen so far for the attractive and increasingly popular suggestion that, by developing the idea of an interpersonal point of view, we get better accounts of both the content and the authority of morality than anything the consequentialists can offer.

ROBERT H. MYERS
York University, Toronto

Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009. Pp. xiv+230. \$35.00 (paper).

Korsgaard's work has been among the very best in ethics in the last twenty-five years, and *Self-Constitution* does not disappoint; it is one of the most interesting books in contemporary ethics. The argument of the book is intricate and difficult, but the rewards of working through its details are very high.

I'll start with a quick and obviously rough outline of some of the book's central moves. Some activities have what Korsgaard calls "constitutive standards." (In fact, Korsgaard suggests that they all do, but I'll leave this aside.) The standards that determine that some houses are better than others are internal to the activity of building a house. They are the constitutive standards of building a house and determined by the function of a house; since a house is supposed to serve as shelter, it is a better or worse house depending on how well it serves this function. According to Korsgaard, once you engage in the activity of building a house, these constitutive standards will apply to you. Korsgaard claims that these constitutive standards wear their normativity on their sleeves; their normativity is simply a consequence of the nature of the activity itself that is what

it is only insofar as it is guided by these standards. But if there are constitutive standards of rational agency as such, constitutive standards that apply to all action insofar as it is action, these standards have normative force for all rational agents. According to Korsgaard, all agency, rational or not, is self-constitution. The giraffe acts in order to preserve its "giraffeness," its actions are all forms of preserving its teleological unity, of preserving itself as a giraffe and preventing its parts from collapsing into scattered bits of matter. Of course, teleological organization does not suffice for action; plants are teleologically organized but do not act. To act is to be "guided by a conception that the animal forms of its environment" (97). The movements of the brutes are determined by a purpose, but the animals need not have a conception of the purpose of their action; they just have a purpose and are guided in their pursuit of this purpose by their representation of the environment. For an owl to act, it need not have any thoughts about the nourishing effects of eating mice; it needs only to perceive mice in its environment and be moved by this perception. However, human beings are guided not only by their conception of the environment but also by their conception of a purpose; given the self-conscious nature of a human agent, any purpose proposed to her by her nature, any instinct or inclination, is one that she may choose not to act on. So a human being can only act on a principle that she identifies with, a principle that she can choose as her principle of action. She must choose an action according to a law that makes an action her action, rather than movements that are caused by some parts of her body. The function of action is, on Korsgaard's view, to unify the agent, to make her the source of her movements, rather than the loose mereological sum of bodily parts that originate disparate movements.

The hypothetical and the categorical imperatives are constitutive norms of rational action, insofar as they are necessary requirements of the activity of determining yourself to be the cause of an end in accordance to principles. An agent can be herself the cause of her movements, only if she is guided by these imperatives. The hypothetical imperative is necessary if I am to be the *cause* of my action and the categorical imperative is necessary if *I* am to be the cause of my action (cf. 72). In other words, the categorical imperative is the principle of autonomy; it constitutes the action as having its source in the agent herself, rather than a movement of some part of the agent. The hypothetical imperative is the principle of efficiency; it constitutes the action as the actual pursuit of the end that the agent sets for herself, rather than random movements. Thus, these principles are normative insofar as without them there is no unified agent and no action.

So, according to Korsgaard, there are constitutive standards of certain activities whose normativity is beyond question; even a bad builder is trying to build an excellent house, but through inattention, incompetence, or other failures, the bad builder builds something that might fall far short of this ideal (39). Of course, Korsgaard recognizes that some shoddy builders are shoddy by design; being dishonest or otherwise corrupt, they build a house for money or other benefits. Such despicable builders might knowingly use cheap materials that will fail to insulate properly or tile the roofs with the aim of shortening their workdays rather than keeping residents dry. But Korsgaard follows Plato here in saying that such builders are not engaged in building a house but rather

in money making or some other activity. (Korsgaard actually seems to waver on this point. She says that “even the most venal and shoddy builder must try to build a good house.” But then two pages later she recognizes that a dishonest builder is not trying to build a good house; she claims there that “he is not trying to build a house at all”; 29, 31.) One is properly said to be building a house, only insofar as one is guided (primarily?) by the function of giving shelter. This kind of move has been widely criticized. Suppose we did find a conceptual connection between an activity counting as “building a house” and the goal of building a good (or perhaps even perfect) house. However, understood this way, the claim that all house building aims at a good house is just a claim about the semantics of “building a house”; it leaves plenty of room for shoddy builders to be engaged in activities such as “schmuinding a house.” What seemed to be an important discovery about the nature of practical rationality turns out to be just an exercise in linguistic stipulation. Barring such definitional fiat, why should we say that corrupt builders are not engaged in building houses? (See, for a recent version of this kind of criticism, David Enoch, “Why Idealize?” *Ethics* 115 [2005]: 759–87.)

However, the focus on the house-building case can prevent us from appreciating the full force of Korsgaard’s argument. At first, it seems that this point will carry over to any attempt to arrive at constitutive standards for agency that can function as something like the foundations for morality; a conceptual connection between agency and any normative standards could deprive those who fail those standards the right to be called “agents.” But compared to the riches and powers that prudent evildoing can bestow someone, the lack of entitlement to the label “agent” seems like a trifle.

Korsgaard argues that the function of an action is to unify (and thus constitute) the agent and that the constitutive norms in question are constitutive of some effect having a source in the agent; that is, they are the conditions under which anything can count as my doing. Thus, first insofar as anything makes a normative claim on the agent, the thing on which it makes a claim must be at the very least a (unified) something to which a normative claim can “attach.” If Korsgaard is right that this kind of thing only exists insofar as it is committed to these imperatives, then agents can act on reasons, and can have reasons to act, only if they are committed to these imperatives. If Korsgaard’s view is correct, all practical normativity depends on the normativity of these imperatives. So, failure to live up to those imperatives doesn’t simply make a concept inapplicable to you, it leaves you choosing nothing, without reasons to do anything, and really being nobody when it comes to acting. Once we see this point, the idea that a “schmagent” could gladly pursue wrongdoing must sound hollow: if Korsgaard is right, there would be no unified thing that could be the subject of an undertaking done for reasons. Of course, one could go on to claim that we can talk about “schmeasons” and “schmunified,” but here the objector owes us an explanation of why we are left with anything that is recognizably similar to the ordinary notion of “action” and “agent” that we originally used to pose the question of how one should act.

However, dismissing the “schmagency objection” this way leaves Korsgaard exposed to another difficulty. How plausible is it to say that those who flout these imperatives are not unified agents at all, that their bodies are the seat of

movement but no (not much?) action? It is worth noting that the view is actually quite plausible with respect to the hypothetical imperative (this by itself raises doubts about the unrestricted application of the “schmagency” objection). Someone who systematically failed to act in accordance with the hypothetical imperative would be hardly recognizable as an agent; in fact, it might not even be coherent to talk about something systematically failing to act in accordance with the hypothetical imperative. And given the impossibility of such systematic failure, it seems plausible that more isolated instances of such failings are cases of defective actions and their agents bad (or imperfect) agents.

But there seems to be no upper limit for evil such that we no longer recognize the subject of evil as an agent or a self. It is to Korsgaard's credit that she faces such issues head on and that a great deal of the book is dedicated to show that, first, the categorical imperative (and also the moral law) is a constitutive principle of agency and that, second, it does provide a compelling account of the nature of evil action. (Because of these special challenges regarding categorical imperatives, I'm setting aside Korsgaard's excellent discussion of the hypothetical imperative.)

Before we assess the cogency of the argument, we should distinguish a few different questions. First, we can ask why are these imperatives, and the categorical imperative in particular, constitutive of agency? Second, if the categorical imperative is a constitutive principle of action, how should we understand the possibility of heteronomous actions? How do they fall short of being a proper action? Finally, does the categorical imperative, understood as the constitutive principle of unified agency, classify as bad all (and only) those actions that we intuitively think are bad?

Let us start with the first issue. Korsgaard's argument starts from her claim that one cannot be a rational agent if one is merely a passive spectator of the various forces battling for control of one's body. If our picture of agency allows only for a battle of appetites in which the winner causes your body to move, it is this appetite, not you, that causes your action. Therefore, this picture leaves no room for action; action must be something you do. But if it is you rather than some part of you that determines your action, then you must identify with your principle of choice; that is, being determined by such a principle must count as self-determining or autonomous. Korsgaard argues that it follows that you must be guided by the categorical imperative; that is, you must will that your principle be a universal law. According to Korsgaard, a will that does not act on universal laws, a “particularist” will, cannot identify itself with the incentive on which it supposedly acts. A particularist will chooses to do X here and now without committing itself to always doing X in similar circumstances or to having a reason to do X in all such circumstances. For instance, it would eat its dessert just this once without endorsing principles such as “when the dessert is particularly good, you should go for it” or “it is permissible to deviate from one's diet on occasion,” and so on. Korsgaard argues that such a willing would not be a proper case of my willing; any attempt to simply identify with a particular incentive would “eradicate the distinction between a person and the incentives on which he acts” (76).

Why must we lose track of the distinction between the particularist willer and her incentive? After all, the incentive was effective only because the agent

chose to act on it in this particular occasion. Moreover, if one's identification with a universal principle of choice can be effective, why not one's identification with a particular incentive? If I can say, "I am the one who chooses to always act on the incentive to be generous," why can't I also say, "I am the one who now chooses to act on the incentive to be generous"? Korsgaard argues that since no general feature of my incentive is relevant for my choice, there is really no content to the thought that I chose to act on this incentive. But this seems to beg the question. The particularist willer can insist that it was her choice to be generous; it's only that her identification of generosity does not extend beyond this particular case. (Sometimes Korsgaard seems to be saying that I'd be unable to pick out such an incentive, but this cannot be what's at issue; being a particularist willer does not prevent me from using general terms to simply refer to what I identify with.)

According to Korsgaard, a particularist cannot commit to act on an incentive; if I commit on Monday to go to the dentist on Tuesday only if I happen to identify with my commitment again on Tuesday, then I committed to nothing at all. I take it that the problem is not that the content of the commitment is conditional (it need not be), but given that the particularist identifies with this incentive only right now, the commitment doesn't have authority over me when Tuesday comes.

But this only shows that the particularist willer can make commitments only if her act of identification can extend beyond the current moment, if when she says, "I identify with the incentive of health here and now," her "now" can extend all the way to Tuesday. If, for instance, the identification is with the action of "getting her teeth cleaned," then if she is a successful agent, her identification with this particular incentive must endure at least until she leaves the dentist.

One might argue that since her identification is not universal, there is little reason to think that her commitment will or should be stable; Tuesday morning she could simply identify with another incentive. But it's unclear that universality could help with this problem. Since, as Korsgaard recognizes, identifying with a universal principle does not preclude changes of mind, the agent could also tomorrow identify with a different universal principle. Unless the universality somehow ties my will to its object, it cannot help. But why would universality do better than particularity at the task of holding on to the same object through time? Universality and particularity are just restrictions on the nature of the object that one is trying to hold on to.

There is an obvious path to try to avoid these problems: one could first say that the incentive chosen must be something that the agent recognizes as a reason and then try to make the case for the universality, or the necessity, of rational commands. However, I suspect that Korsgaard avoids this route because she wants the constraints on what can count as a reason to flow from constraints on self-constitution rather than the other way around. Although this creates problems for her view that I am not sure can be overcome, it does bring to light some of its advantages too. Korsgaard's view can explain why, in a Kantian view, human beings ought to follow the categorical imperative even when the categorical imperative commands them to act against strong inclinations. The claim that the categorical imperative expresses the voice of reason might ring hollow in this context; after all, the question is exactly why should I listen to the voice

of reason rather than to the appealing chants of inclinations. Korsgaard convincingly argues that the Combat Model, the model that passion and reason battle for control of my actions, neither answers this question nor explains why it counts as my action when either force wins. On Korsgaard's account, reason is to be heard not because of an unexplained preference for this force within the whole but rather because only when reason is heard, does it makes sense to say that there is a whole and that I have acted. The agent acts when she is unified by the rule of reason for the good of the whole, just as a government acts only when its actions are determined by the proper powers and procedures.

Given that the categorical imperative is constitutive of agency, Korsgaard is committed to the view that heteronomous actions, actions that are not guided by the categorical imperative, are defective qua actions. Korsgaard relies on an interpretation of Plato's progressively more unjust (and disunified) constitutions in order to explain why these actions are defective to varying degrees. Korsgaard makes a persuasive case that the agent who always follows the strongest passion of the moment (the democratic constitution) is not a fully unified agent. However, it is hard to see why the other "bad" constitutions are cases of defective agency; one wishes for a better understanding of why the prudent, honorable, or tyrannical (single desire) constitution is incapable of unifying the agent. After all, Korsgaard grants that "honor and prudence are principles of choice sufficiently like true virtue to hold a soul together through most kinds of stress" (166). Even ignoring for the moment that it is unclear what kind of "stress" we are talking about, we now run the danger of having illegitimately turned unification into a goal to be maximized. We are no longer talking here about perfect or imperfect unification but rather about how stable unification is and its comparative capacity to persist. But even if we grant that the just constitution will be unified in more possible situations, it does not seem to follow that it is the only one that properly acts when it acts. A crystal chalice is more likely to break and lose its unity than a glass one, but it's not a defective chalice for that reason.

In fairness to Korsgaard, she does not think that her defense of the categorical imperative on its own can show the normativity of morality. For I can accept that universal principles prescribe certain reasons for action but think that the reasons they prescribe are "private" reasons. (This is similar to her distinction in *Sources of Normativity* between the categorical imperative and the moral law.) I can accept that it is a universal law that one should always pursue artistic excellence but think that this gives rise to a reason for me to pursue my artistic excellence and gives you reason to pursue your artistic excellence but to no reason that we share. According to Korsgaard, in order to get from the categorical imperative to the moral law, we need the further assumption that all reasons are public; that is, that if there is a reason for me to pursue my (or anyone's) artistic excellence, then there is simply a reason, valid for everyone, to pursue my (or anyone's) artistic excellence, a reason that is normative to all of us. I can only sketch here Korsgaard's complex argument for the view that reasons are public. The argument takes as its starting point two competing views of interaction, of what happens when you and I are trying to do something together. On one view, I treat your actions and also your reasons as just part of my circumstances, something that I have to enter in my instrumental calculation of how I will act given my reasons. On the second, Kantian view, in interactions

we treat each other's reasons as a reason, so that when I recognize that you have a reason not to have our meetings in a crowded coffeehouse, I take this to be a reason for us not to meet at a crowded coffeehouse. If we extend this point to all our interactions, we see that, on the second view, reasons are public. My reason to take the elevator gives everyone a reason to act, though of course the reason will call for different agents to perform different actions; it will call on me to get into the elevator and for you, perhaps, to open the door for me. One moment of reflection on what the first view implies about, for instance, conducting a conversation with a friend should be enough to see how implausible the first view is in its full generality. But whether this shows that the second model of interaction is correct depends on whether these two understandings of interaction are exhaustive and also on whether we need to treat all interactions on the same model. Korsgaard is aware that this is what she needs to show, but I am not completely confident that she succeeds. Here I can't go over all the possibilities, but I'll just look at one she considers: that I treat other people's reasons as reasons because I care for them. This would seem to have at least one important advantage: it allows that I do not treat everyone's reasons as reasons or at least not in the same way or to the same extent; my reasons or the reasons of my loved ones might be privileged for me. But all that Korsgaard says against the view is that "it doesn't work that way: it is not that I decide that because I love or respect someone I will accord normative force to his reasons. It's rather that finding myself loving or respecting someone, I simply do accord normative force to his reasons" (202). I am not sure I understand why Korsgaard thinks this is true or why she thinks it poses a problem for the view in question. Korsgaard's opponent might be committed to thinking that "I treat Sally's reasons as reasons because I care about her" is an acceptable, final explanation for treating Sally's reasons as reasons for me. But why is this explanation incompatible with "simply accord(ing) normative force" to the reasons of my loved ones? Even if one simply "finds oneself according normative force to the reasons of those one cares about," why would it be incoherent to extend the normative force of those reasons only to those one loves and perhaps even proportionally to one's love? This would seem to be a form of universal legislation that has not been yet ruled out.

Of course, I cannot here discuss all the arguments in this rich book. Excellent parts not covered include her argument against Sidgwick's view on the indefensibility of our ordinary attitudes to lying and the details of her insightful discussion of the nature of teleologically organized beings. This is a book that no serious ethicist can afford not to read carefully in its entirety. Moreover, the criticisms raised here are not supposed to be conclusive. They are only attempts to bring to focus the places where we might still need a few more brushstrokes in this extremely compelling and original picture of rational agency.

SERGIO TENENBAUM
University of Toronto