

© 2011 The Author

Metaphilosophy © 2011 Metaphilosophy LLC and Blackwell Publishing Ltd

Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK, and
350 Main Street, Malden, MA 02148, USA

METAPHILOSOPHY

Vol. 42, No. 4, July 2011

0026-1068

KORSGAARD'S REJECTION OF CONSEQUENTIALISM

DAVID CUMMISKEY

Abstract: In her recent book *Self-Constitution: Agency, Identity, and Integrity*, Christine Korsgaard does a wonderful job developing her Kantian account of normativity and the rational necessity of morality. Korsgaard's account of normativity, however, has received its fair share of attention. In this discussion, the focus is on the resulting moral theory and, in particular, on Korsgaard's reason for rejecting consequentialist moral theories. The article suggests that we assume that Korsgaard's vindication of Kantian rationalism is successful and ask whether, nonetheless, her account is consistent with consequentialism. It suggests further that we grant that moral reasons are not based on substantive principles, and that they must instead emerge from the purely formal principles of practical reason. Can consequentialist principles nonetheless emerge from the formal constraints of practical reason? Why can't a consequentialist embrace Korsgaard's account of self-constitution and normativity?

Keywords: Korsgaard, self-constitution, integrity, Kantian ethics, rationalism, consequentialism.

For more than two decades, Christine Korsgaard has been developing a remarkably clear and comprehensive account of Kant's moral theory. In addition to her excellent interpretive essays on almost every aspect of Kant's ethics, she has defended a systematic Kantian account of the basis of morality, the nature of normativity, and the relationship of normativity to agency, rationality, and autonomy. Her recent book *Self-Constitution: Agency, Identity, and Integrity* (Korsgaard 2009) picks up on many of the themes of *The Sources of Normativity* (Korsgaard 1996) but also presents her current views with exceptional clarity and force. In a candid example, near the end of *Self-Constitution* Korsgaard suggests that part of her practical identity, one of her personal goals, is to write a book on Kant good enough to be required reading in all ethics classes. She may well have succeeded in this personal ambition: *Self-Constitution* should indeed be required reading.

When it comes to defending the categorical nature of moral imperatives, Korsgaard has done more than any other contemporary philoso-

© 2011 The Author

Metaphilosophy © 2011 Metaphilosophy LLC and Blackwell Publishing Ltd

pher.¹ I agree that only an account like hers can capture the categorical nature of morality. Her account of normativity, however, has received its fair share of attention. My focus is on the resulting moral theory. Just as it is an open question whether or not Rawls's contract argument justifies his two principles of justice, as opposed to some version of consequentialism, so too it is an open question whether Korsgaard's account of the rational necessity of the Moral Law is incompatible with a consequentialist interpretation of the Moral Law. I shall assume that Korsgaard's vindication of Kantian rationalism is successful and ask whether it is consistent with normative consequentialism. Why can't a unified agent nonetheless be a consequentialist? Why can't a consequentialist embrace Korsgaard's account of self-constitution and normativity?

Korsgaard's Argument

First, Korsgaard rejects the common conception of action where the function of actions is to bring something about. This is Mill's view, for example, and it leads him much of the way to utilitarianism. Korsgaard calls this familiar idea the "production conception of action," and she grants that it naturally leads to the idea that actions should be judged by their consequences, and this in turn implies that the deontologist must justify any constraints on the maximization of the good (2009, 8–9). As alternative to the "production conception of actions," she argues that actions are composed of both the act and the purpose, where the act is done for the purpose (although the purpose may involve doing the act for its own sake). The evaluation and justification of actions depend on the relationship between the act and the purpose. The maxim of an action spells out the act and the purpose, and the action is justified when doing the act for that purpose constitutes a good reason. Actions are intelligible objects that embody reasons (Korsgaard 2009, 14).

It follows that there is no simple route from a conception of action as production to consequentialism. If actions embody reasons, the nature of the normativity of reasons remains an open question, and so consequentialism is not the default position. As a corollary, if one rejects the production view of action, deontological constraints are not puzzling or paradoxical; if certain actions cannot embody reasons, these actions are intrinsically impermissible, or just plain wrong, and that's that. The dispute between deontology and consequentialism is thus a dispute about the nature of (moral) reasons.

¹ Although Rawls resurrected many Kantian ideas, unlike Korsgaard he largely abandoned Kant's rationalist project. Instead of a *Grounding of the Metaphysics of Morals*, we get *Political Liberalism*, which is grounded only on contemporary Western political ideals—a political, not metaphysical, theory of justice. In contrast, Korsgaard truly strives to understand, articulate, and vindicate Kant's project of capturing the rational necessity of morality.

Second, Korsgaard argues that our actions constitute our agency. They reflect the practical identity of particular agents. Certain actions (that is, acts done for a purpose) are required by particular contingent identities, and in acting from that practical identity we constitute ourselves (Korsgaard 2009, 20–21). The normativity of the actions is rooted in our practical identity, and we affirm our contingent identity through the action. As she explains it, “valuing yourself under a certain description consists in endorsing the reasons and obligation to which that way of identifying yourself gives rise” (2009, 24). The unity of the persons and the overall normativity of actions are determined by the coherence of our diverse contingent identities, and also, as we shall see, the test of the Hypothetical Imperative and the Categorical Imperative.

It is important to note that Korsgaard does not embrace an unduly voluntarist, disembodied, or unencumbered conception of practical identity. She grants that some practical identities are initially unchosen. Our embodied, finite animal nature is an essential part of our human nature. We are also, of course, born into our familial and cultural identities, and we often simply find ourselves in certain relationships and roles. Of course, other identities, like one’s profession, are reflectively chosen and cultivated. The important point for Korsgaard is that all of our practical identities are normative, and constitutive of us, when we endorse and reaffirm them through our particular actions.

Third, the justification of actions must involve the form of the maxim of the action—and not a substantive principle of reason. This is indeed the core point of all of Korsgaard’s arguments. (As Korsgaard states, “I am *always* making the same argument” [2009, 76].) *Moral principles must be formal principles of reason.* The two formal principles of Kantian reason are familiar. The first is the Hypothetical Imperative (HI): action is simply purposeful determination. Determination is a form of causality, so action involves a form of causality. If you will an end, you determine yourself to be a cause of it. The cause is a means to the willed end, and so if you will the end, you also will some necessary means to the end. In short, willing the means to your ends is a necessary requirement of reason (Korsgaard 2009, 72–73). The HI is thus a purely formal principle that is constitutive of reason itself.

So far, so good. Next, of course, we have the Categorical Imperative (CI); Korsgaard emphasizes our ability to step back from our contingent identities and reflectively endorse or reject them. When we deliberate and act for a purpose, we necessarily take ourselves to be something *over and above* the particular incentives and parts of our complex practical identities. Korsgaard argues that reasons must be universal because otherwise there is *nothing over and above* the various incentives (2009, 75). She calls this the “argument against particularistic willing” (2009, 72). The upshot of this argument is that particularistic willing isn’t willing at all, and thus willing involves taking one’s maxims to be at least provisionally universalizable.

In making this argument, Korsgaard distinguishes general principles from universal principles and provisionally universal principles (2009, 73–75). She emphasizes that universality is not generality; our principles can be finely tuned, and they should be sensitive to the particulars of our situation. In addition, our principles will often be *provisionally universal* in that we recognize that we might not have “thought of everything,” and thus our principles are (always) revisable in light of new information or insights (Korsgaard 2009, 73). Korsgaard explains that “particularistic willing would be a matter of willing a maxim for exactly this occasion without taking it to have any other implications of any kind for any other occasion” (2009, 75). The problem here is that “self-determination requires identification with the principle of choice. . . . Particularistic willing eradicates the distinction between the person and the incentives within him. And then there is nothing left that is the *person*, the agent, that is his self-determined will as distinct from the play of incentives within him. . . . Particularistic willing lacks a subject, a person who is the cause of his actions. So particularistic willing is not willing at all” (Korsgaard 2009, 76).

Since willing must take the form of universalizable maxims, the CI is a constitutive and formal principle of willing itself. Indeed, self-determination presupposes the CI, and thus the CI is simply the principle of the autonomy of the will.

It follows that, unlike our contingent practical identities, we do not choose the HI or the CI, for these principles just are the form of willing itself. The HI is the form of willing as determination and causality, and the CI is the form of willing insofar as the causality is one’s own. The HI is the form of efficacy, and the CI is the form of autonomy (Korsgaard 2009, 131).

Fourth, according to Korsgaard, there is still another step to get to what she calls the Moral Law. “The Categorical Imperative is the law of acting only on maxims that you can will to be universal laws. The Moral Law, as I characterized it (in *The Sources of Normativity*), is the law of acting only on maxims that all rational beings could act on together in a workable cooperative system” (Korsgaard 2009, 80). When we universalize our maxims, they must be endorsable as public reasons for all rational beings. They must be public and shareable. Although our practical identity is largely contingent, our identity as a rational (end-setting) self is constitutive and necessary. We cannot deliberate, act for a purpose, and escape this identity. All action presupposes the normativity of our rational nature (Korsgaard 2009, 24). It thus follows that we must see ourselves as members of a Kingdom of Ends (Korsgaard 2009, 41).

The Rejection of Consequentialism

Let’s assume that this is all correct. How does it exclude consequentialist normative principles? Why can’t the principle unifying and constituting the self, as a member of a Kingdom of Ends, be consequentialist nonetheless?

The first issue here is whether all consequentialist principles are committed to a substantive conception of moral principles, and thus violate the CI, which is the principle of the autonomy of the will. On the formal conception, moral principles are tests of maxims. This means that moral reasons just are the reasons that emerge from the correct process of deliberation, and they are binding because they are the conclusion of our deliberation (Korsgaard 2009, 49). So we can conclude that consequentialist principles are binding moral principles if they can emerge from the appropriate process of deliberation.

Korsgaard argues, however, that “the maximization of satisfaction over time, or any other form of overall good, is a *substantive* end, not a formal one, and any principle directing us to promote and prefer it would be a substantive principle” (2009, 53). Now, Korsgaard’s point is that we cannot identify the principle of prudence or desire satisfaction with the *formal* principle of the HI.² Any maximizing principle cannot assume or presuppose a substantive end *that is prior to the formal principles of deliberative rationality*. Instead, moral ends must be necessary ends that emerge from the correct process of deliberation itself. From this argument, it follows that the ends promoted by a consequentialist moral principle must be ends that pass the HI and CI tests for correct moral reasoning, or that emerge as required by moral reasoning. The starting point for any Kantian form of consequentialism will be necessary ends of agency; that is, ends that are based on pure practical reason alone. Of course, Kant thought that our own perfection and the happiness of others were just such ends, and so the argument thus far actually does not rule out, in principle, that there are necessary ends that we should promote as effectively as possible. Of course, Kant also thought that these ends are captured by imperfect duties that are limited by perfect duties, but this is exactly what needs to be shown and not just assumed (see Cummiskey 1996). It thus follows that morality could require “the maximization of satisfaction over time, or any other form of overall good,” provided that the conception of the good emerged from correct moral reasoning.

In addition, Korsgaard favors the “practical contradiction” interpretation of the formula of universalizability. The basic idea here is that universalizing the maxim of action would undermine my ability to advance my own ends by acting on the universalized maxims. On the practical interpretation, why can’t a consequentialist will that everyone accepts the favored consequentialist principle of right action? How does

² This raises an important problem for Korsgaard’s account of self-constitution, because the HI does not provide any direction as to how to unify one’s ends into a coherent self (it is simply a principle of efficacy), and the CI excludes impermissible maxims but does not require any particular unity in the incentives that pass the universalizability test. So without what she dubs “the missing principle” it is not clear how we get the supposed requirement for unity and integrity in the self (Korsgaard 2009, 53). It is also not clear how reason can guide deliberation that is evaluating competing permissible ends.

universalizing a consequentialist principle of rightness result in a practical contradiction? I don't see why it would. If everyone is a consequentialist, this would further promote the good, not undermine it.

This brings us to the next issue. This second issue is whether consequentialist principles violate the public nature of reasons. On its face there is nothing in Korsgaard's account of the public nature of reasons that conflicts with consequentialism. Indeed, her rejection of the distinction between agent-relative and agent-neutral reasons, in her response to Williams-based concerns, mirrors Peter Railton's response in his defense of consequentialism (Railton 1984). Both Korsgaard and Railton argue that an agent's personal reasons are taken to have interpersonal significance, and that there is a problem with the idea of private reasons that is similar to the problems with a private language.

Korsgaard might object that consequentialism violates the "publicity condition"—that is, that moral reasons must be promulgated and publicly defensible. In principle, although not in practice, the best outcomes might be promoted if people were not consequentialist, and so in principle at least consequentialism could be self-effacing. The reason for this is that consequentialists emphasize the distinction between the right-making characteristics of actions and the decision-making procedures of agents. (This distinction is a corollary to the distinction between the truth conditions of a theory and the acceptance conditions.) Should a Kantian reject this distinction? On the one hand, the HI/CI test is just the form of good reason, and thus it is simply the form of justified reasoning. So the decision-making procedures and standards of rightness collapse here in an interesting way. On the other hand, the HI/CI test of right actions surely would yield the conclusion that agents should not always deliberate in accordance with the CI procedure; and so Kantian reasoning also would be partly self-effacing. More radically, in principle, could the CI permit or require its total self-effacement too? Hypothetically, if a demon threatens to destroy all of humanity unless we all consent to take a drug (it could be added to the atmosphere) that makes us all accept substantive principles of morality (but that in other respects fully treats humanity as an end-in-itself), would the CI permit or require this action? I find this question open and interesting.

In addition, recall that the formal CI test is not the Moral Law. The CI prohibits particularistic willing, and there is absolutely no reason to think that consequentialism would require particularistic willing. Since consequentialist decision-making principles would be provisionally universalizable, why would consequentialism violate the purely formal CI test?

A closely related third issue is whether consequentialist principles recognize the necessary priority of truth in the requisite sense. Korsgaard approaches this interesting question in her own distinctive way. Instead of focusing on the utility of deceiving others, she focuses on the application of the principle of utility to self-deception. (Her Platinum Rule is at play here: "Do unto others as you cannot help to do unto yourself.") On the

utilitarian principle, at the most basic level, I should subordinate concern for the truth for its own sake to the utility of truth. Korsgaard asks, how is this supposed to work in my own (first-person) case? I must at least tell myself the truth about whether the truth is useful or not, and so, in the first-person case, truth must be prior to utility. The principle of utility cannot serve as a basic unifying principle for oneself, and this shows that it is also unsuitable as an interpersonal principle (Korsgaard 2009, 182–83).

In response, again, we might start by distinguishing the standard of right and the decision-making procedure. Yes, this is so as far as the decision procedure goes. We need to know the truth about *the utility of the truth* to decide whether truth should be always basic and prior to considerations of utility. This does not show that truth-telling is a basic right-making characteristic.³ Indeed, for Korsgaard the HI and CI are basic; and so *the truth is prior to utility* only if the CI and HI show that it is. But, of course, this means that we must know the truth about whether the principle of the priority of truth is efficacious and universalizable, so does this mean that the HI and the CI are subordinate to truth? Is this not a general problem for any theory? How does the Kantian approach avoid this problem? Again, I think that there is much that both sides can say here, and it seems worthy of additional exploration.

It is also noteworthy that recent empirical work on memory and personal integrity suggests that selective forgetfulness plays a central role in the construction of a practical identity.⁴ The enhanced memory capacity of hypermnesia—that is, remembering everything about oneself—is actually disabling in interesting ways. It seems that an enhanced memory can undermine selective retention, and that as a result one loses the ability to construct a coherent sense of self. This implies that the process of self-constitution subordinates truth in a way that facilitates the self-construction of a coherent self-narrative. If this is so, selective retention may actually be required by the efficacy demand of the HI. Since the HI and the CI are the form of practical reasoning, the practical role of truth must be decided by good reasoning; to think otherwise would

³ In addition, even if first-order truth is basic (start with truth), this does not show that the truth is not defeasible second-order—say, lying to prevent lying or to protect the need for first-order truth. Even if coercion and deception are pro tanto first-order duties, they may not be second-order constraints. This will depend on whether the reasons we can share permit these kinds of deceptions or not.

⁴ The case I have in mind is that of Jill Price, who has hyperthymestic syndrome, or autobiographical memory syndrome (see Price 2008 and Parker et al. 2006, the study by the neuroscientists who examined her). Price writes, “Whereas people generally create narratives of their lives that are fashioned by a process of selective remembering and an enormous amount of forgetting and continuously re-crafting that narrative through the course of life, I have not been able to do so” (2008, 6); “I’ve come to understand that there is real value in being able to forget a good deal about our lives” (45). I thank Walter Glannon for drawing my attention to this issue.

be to treat truth as a substantive principle of reason. The autonomy of reason, not empirical truth, is the sovereign source of practical reason.⁵

Conclusion

I have argued that it is an open question whether some form of consequentialism is compatible with Korsgaard's Kantian conception of morality and normativity. Before ending, however, I need to emphasize that the rejection of consequentialism is not the topic or focus of *Self-Constitution*. Korsgaard's focus is on the nature of integrity and agency, and a real strength of her book is its comprehensive account of the nature of *defective* agency—indeed, of all of the ways that we, as agents, can go wrong. I have focused on the Kantian elements of her position, but *Self-Constitution* also incorporates core insights from Aristotle and Plato. Finally, in addition to her account of embodied human agency, Korsgaard develops a complementary account of animal agency. Although the rejection of consequentialism is not Korsgaard's focus in this book on agency, identity, and integrity, I have tried, first, to spell out as clearly as possible her reasons for rejecting consequentialism and, second, to explore the possibility that a consequentialist could accept her account of agency and normativity.

Philosophy Department
Bates College
Lewiston, ME 04240
USA
dcummisk@bates.edu

References

- Cummiskey, David. 1996. *Kantian Consequentialism*. New York: Oxford University Press.
- Korsgaard, Christine M. 1996. *The Sources of Normativity*. Edited by Onora O'Neill. Cambridge: Cambridge University Press.
- . 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Parker, Elizabeth, et al. 2006. "Case of Unusual Autobiographical Remembering." *Neurocase* 12:35–49.
- Price, Jill. 2008. *The Woman Who Can't Forget*. New York: Free Press.
- Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs* 13, no. 2:134–71.

⁵ Consequentialist principles also justify coercion and deception when necessary to advance the good. The consequentialist thus asks the Kantian, why can't a dutiful agent consent to being coerced or deceived when doing so is necessary to promote the good? Of course, I cannot consent at the moment that I am coerced or deceived, but this is also true of all justified coercion and law enforcement. If consent to a system of coercive laws is consistent with the Moral Law and treating persons as ends, why can't we also consent to coercion justified by consequentialist considerations?