

Review of *Self-Constitution: Agency, Identity, and Integrity*. By CHRISTINE M. KORSGAARD. (Oxford: Oxford University Press, 2009. Pp. xiv + 230. Price £45.00.)

Forthcoming in *Philosophical Quarterly*

This is the author's copy, which may differ from the final print version

In *Self-Constitution*, Christine Korsgaard continues to pursue a neo-Kantian project that attempts to ground morality in constitutive features of human agency and the self. She has defended a version of this view in her book *The Sources of Normativity* (1996, Cambridge University Press). As Korsgaard points out in the first chapter, *Self-Constitution* has the same conclusion as *Sources*—that is, roughly, the claim that in order to act and constitute oneself *well*, one must act in accordance with Kant's categorical imperative and value the humanity in others and oneself. And like *Sources*, this new book is based on a series of lectures (the Locke lectures given at Oxford in 2002).

Not only the conclusion, but also many of the themes and arguments in *Self-Constitution* will be familiar to anyone who knows Korsgaard's work. Despite that, the book has many new ideas and arguments on offer. Korsgaard says that what distinguishes this book is a direct focus on the nature of human agency (p. 25). This is clearly an important contribution to the overall argument, which says that the normativity of morality has its source in essential features of agency and the agent. But what I found to be the most interesting aspect, and the real innovation of *Self-Constitution*, is the idea that a defence of Kantian ethics can be strengthened by an appeal to Plato's account of the soul (or self). This suggestion is based on the interesting observation that Kant's and Plato's account of the self share some of their core features. Overall, the book is engaging, written in an accessible style, and it makes an important contribution to Korsgaard's neo-Kantian project.

But *Self-Constitution* is also full of controversial claims and arguments. Among them are the claim that the proper goal or task of moral agency is not to be good or rational but to unify the self (section 1.1.5); that one is morally responsible only if one constitutes or creates oneself as an agent (1.4.6); that only autonomous action in accordance with Kant's categorical imperative constitutes and unifies the self well (4.4, 9.7 and 10.4); that the *function* of action and practical reason is to constitute and unify the self (5.1); that the self *must* be unified in processes of self-

constitution because self-consciousness continually divides the self (6.4); that it is up to oneself to *decide* what counts as a reason (6.4.3); that we must decide *who we are* (1.4.3 and 6.4); and that respect for humanity is a necessary condition for effective agency (9.7).

I suspect that many of those claims will provoke an immediate negative reaction with most philosophers, including Kant scholars. Some of those claims seem so strong that one cannot help thinking that something must be wrong with the arguments for them. In the light of this, one can only admire Korsgaard's persistence and ingenuity in the pursuit and defence of her view. And the focus on agency and Plato's conception of the self certainly adds some new twists to the story. In this short discussion, I shall set aside the more familiar worries concerning Kantian ethics. I will instead focus on some of the novel contributions that *Self-Constitution* has to offer.

As Korsgaard explains (7.1), Plato's account of the self provides a resolution to the following dilemma. If one construes the self, or the agent, as a mere bundle of mental states, one concedes, in effect, that there is no self. All there is on this view, according to Korsgaard, is a bunch of forces that are acting on or in "the agent" (which is not really a human agent or person, but a mere animal or organism). But if we assume, on the other hand, that the self is a separate entity, distinct from all the mental events and dispositions, then there are no properties that distinguish it as an individual entity. On this view, the self is a mere substance without an essence. Plato's view offers an alternative in analogy with a constitutional model of the city (state or nation). The city is neither the sum of its parts or inhabitants, nor some abstract entity. Rather, the city is to be identified with the constitutive principles that give it its unity to act as one city. By analogy, one can avoid the mentioned dilemma if one identifies the self with the principles that bind its mental parts together and unify them into a single agent (7.2).

A first thing to note here is that proponents of a neo-Humean approach might very well agree that the self is not a mere bundle, but a holistic system that is organized into a psychological unit. The real controversy comes with the claim that any bundle of mental states and dispositions must be organized in accordance with Kant's categorical imperative in order to constitute a well-functioning and well-unified agent. In the main part of the argument for this claim, which is a central thesis of the book, Korsgaard argues that what she calls 'particularistic willing' is impossible (4.4). And from this she draws the conclusion that willing must be universal (4.4.4). Concerning this last step, it has not become clear to me why the impossibility of particularistic willing entails that willing must be universal. Korsgaard points out that what she

has in mind is ‘absolute universality’ and not mere generality (4.4.2). But there is no clear argument for the claim that willing cannot be merely general (that is, neither particularistic nor universal). But my main worry concerns her argument against particularistic willing. Like other arguments in the book, it is based on rather controversial assumptions concerning the nature of agency. According to Korsgaard, if you have conflicting desires and you choose to pursue one of them, then you *must* identify with the principle of choice that you *must* deploy in making the choice. For if the choice had been settled merely by the strength of your desires, you would be a ‘passive spectator’ and not an agent (4.4.3). However, proponents of a Humean or event-causal account of agency would reject—quite rightly, I think—the assumptions about agency that underlie this argument as question-begging.

Herein lies, in my opinion, a major shortcoming of the book. Korsgaard does provide arguments for her claims concerning the sources of normativity. But many of them are based on strong assumptions concerning the nature of human agency that are themselves in need of justification. And the problem is that few of them are properly justified in this book. Korsgaard asserts several times that a movement can be attributed to an agent as an action only if it is ‘an expression of the agent as a whole’ and only if it is not the outcome of ‘forces working on or in her’ (1.4.1, 4.4.3, 5.3.1 and 5.5). Concerning the first claim, it is not obvious that there is no sense in which this can be true on a Humean account. In other words, it remains to be shown that a Humean or event-causal theory of agency cannot give *some* account of agency that is the expression of the agent as a whole (what comes to mind here is a Frankfurt-style account of wholehearted agency). The second claim seems to beg the question. On the Humean approach, there is a sense in which actions are just that: the effects of forces working on or in the agent—although the Humean would not put it quite like that, of course.

It might be helpful to compare this to the dispute between event-causalists and agent-causalists in the debate on free will. Roderick Chisholm, an influential agent-causalist, pointed out that the controversy is not about whether or not agents are the causes of actions. It is about the question of whether or not the agent-causal relation can be *reduced* to event-causal relations. By stipulating that actions cannot be the causal effects of agent-involving events (“forces that work on or in me”) Korsgaard would settle that dispute *by stipulation*. (Although her own account differs significantly from the view of Chisholm and other agent-causalists.)

The ideal of agency, Korsgaard says, is the ideal of ‘inserting yourself into the causal order’ (5.2.7). Again, opponents can say that there might well be an interpretation of this claim that is true even on Humean accounts of agency. In its literal reading, however, they would reject it not only as begging the question, but as a misguided starting point for an account of what it is to interact with the world. They would insist that there is no need to *insert* ourselves into the causal order, because we are already and always part of it.

Another new and distinctive claim of *Self-Constitution* is that we must constitute and unify ourselves in agency as our self-consciousness continually divides the self. Korsgaard notes that this gives rise to an apparent paradox: ‘how can you constitute yourself, create yourself, unless you are already there?’ (1.4.3) In response she argues that there is no paradox, as the self in question does not exist prior to the choice of the action. Rather, this self is constituted by the choice. But this response overlooks a problem that arises in connection with claims about responsibility. Some philosophers, including Korsgaard, think that we are truly responsible only if we literally create ourselves (1.4.3). As Korsgaard shows, the only way to make sense of self-constitution is to assume that the self is constituted *by the choice*. But if the self is constituted by that very choice, how can it be responsible for it? This, it seems to me, is the real paradox of self-constitution, which Korsgaard does not address (compare 2.4).

Finally, I would like to address one worry concerning the overall project. The crucial starting point, as I see it, is the assumption that ‘the normativity of obligation is, among other things, a psychological force’ (1.1.2). The central idea is that the relevant psychological necessities stem from the structural and constitutive features of agency and selfhood. But the suggestion that the force of normativity derives from the *psychological necessity* to unify the agent strikes me as problematic. To be sure, the view does not derive an “ought” from a contingent “is”, as the source of normativity consists in constitutive and necessary conditions. Nevertheless, it seems that the view identifies the wrong reasons—the wrong *type* of ground. This becomes clearest when Korsgaard agrees with Plato that it is best to have an “aristocratic soul”, because only a soul with this type of constitution is well-unified (8.5). But when we ask, for instance, why one should not be a tyrant (why one should not have a “tyrannical soul”), then an appeal to the unity of the self seems to provide the wrong type of consideration.