

Instrumental Rationality and the Argument Against Particularistic Willing

Christine Korsgaard has defended an argument against particularistic willing in a number of publications. (I focus on (Korsgaard, 2009).) Her core idea is that we cannot will particularistically, so we must will universally, and then the categorical imperative (*CI*) is a constitutive principle of action. She goes wrong, but a better version of her argument shows that instrumental rationality is constitutive of agency, not *CI*.

I begin by presenting my reconstruction of Korsgaard's argument. The core idea is that to act, one has to make oneself into a cause in the world. This requires unifying oneself diachronically by forming a maxim – a representation of the ends and means in an act-type – that is treated as universal. Hence, one is supposed to hold on to it for all future, at least all else being equal. If one does not, it becomes a mere causal factor within one, so the agent does not stand behind the action anymore.

If one universalizes maxims in this way, one universalizes them in accordance with *CI*, for *CI* demands just this kind of universalization. Hence, Korsgaard thinks, *CI* is one's constitutive principle, i.e. what constitutes one as a specific kind of creature (in this case, a human agent). So to act, one must constitute oneself in accordance with the categorical imperative.

I proceed by criticising the alleged need to universalize one's maxims diachronically. First, it seems like one does not need universalized maxims for diachronic unification (cf. Millgram, 2011). One can plausibly unify oneself diachronically without universalizing a maxim for all future, e.g. by forming long-standing intentions.

More importantly, I argue that one does not need to unify maxims – or oneself – diachronically to act at all. We can easily think of actions from someone who is an agent but does not have such maxims – universalized maxims seem theoretically redundant to describe actions. So we do not need them.

Having criticised Korsgaard's universalization assumption, I reformulate her argument. It then looks like this:

- (1) If an agent is to perform an action, then she must make herself into a cause in the world.
- (2) If an agent must make herself into a cause in the world, then the agent must unify herself.
- (3) If an agent unifies herself, then she must control via the instrumental principle.
-
- (C1) If an agent is to perform an action, she must control it by the instrumental principle (from 1, 2, and 3*).
- (5) All properties that are needed for an agent to act are constitutive of agency.
-
- (C2) If an agent is to perform an action, the instrumental principle is constitutive of agency (from (C1) and (5)).

To defend this argument, I first agree with Korsgaard that actions must have causal impact on the world, and involve taking means to ends.¹ And we share premises (1) and (2): One needs

¹ Or, at least, the vast majority of actions work like this.

to, first, make oneself into a cause, and second, unify oneself to act. The first premise solves the problem of the disappearing agent for causal theories of action – it shows how the agent actively contributes something to the causal order via the process of making herself into a cause. The second premise is defensible because action requires that one unifies representations of means and of ends to initiate it. They are distinct states absent some unifying manoeuvre.

To defend (3), I argue that there is a kind of unification involved in uniting one's representations of means and ends if one is steered by a principle of instrumental rationality.² The idea is that these representations – of what one aims to do, and how one can achieve it – are distinct, and so some third factor is required to unify them so that one can take the means to one's end. That third factor is the principle.³ One needs to unify oneself in one sense to be able to initiate and control one's action, and one needs to do it via the principle because one's representations do not do that by themselves (or so I argue, appealing to standard understandings of such mental states). This kind of unification avoids the problems for Korsgaard's form of universalization. One does not need universalized or diachronic maxims to act. One only needs to combine one's present mental states. (C1) follows automatically.

Moreover, if something is necessary to perform an action, it is constitutive of that kind of agency, because agency consists in whatever it is that one needs to act. This gets us (5).⁴ (C2) follows.

However, this does not by itself establish that instrumental rationality is normative or defensible. Since my argument is constitutivist, a full defence would involve responding to objections to that view. But that is another project.

References

Millgram, E. 2011. Review of 'Self-Constitution: Agency, Identity, and Integrity The Constitution of Agency: Essays on Practical Reason and Moral Psychology' by Christine Korsgaard. *Australasian Journal of Philosophy*. 89(3), pp. 549-556.

Korsgaard, C.M. 2009. *Self-Constitution*. Oxford, UK: Oxford University Press.

² It does not matter much how one formulates the principle, if one wants to narrow-scope or wide-scope, etc.

³ This is possible on different ontological interpretations of the principle. It might be a disposition, and then it does causal work as part of the intention-formation process. Or it might be propositional and represented in the agent's mental states, but then the mental states, when held in the right way, cause the unity. That is why I use the generic formulation 'steered by' in the argument.

⁴ It also shows why agency is cheap, and that there therefore is no problem to say that instrumental rationality – even though it is a weak principle – is what contributes positively to show that the action comes from the agent.