

The definitive version of this article is available in *Philosophical Quarterly*, vol. 61:243, 343-364, April 2011 (<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9213.2010.680.x/abstract>)

Korsgaard's Constitutive Arguments and the Principles of Practical Reason

Ariela Tubert

Constitutive arguments for the normativity of rational requirements have received quite a bit of attention lately.¹ The basic idea is that there are certain features that are constitutive of belief or action such that insofar as we believe or act at all, we commit ourselves either to following principles like the law of non-contradiction or the instrumental principle, or to aiming at something like the truth or the good.² Kant is the historical figure most associated with such arguments but the recent interest in them is not limited to contemporary Kantians narrowly defined. For some, the appeal of constitutive arguments is that they promise to explain the normative force of certain rational requirements in a broadly naturalistic way, without appealing to irreducible normative properties external to the agent.³ For others, the appeal is that they promise to explain how rational requirements can get a hold on the agent, leaving no room for a skeptical challenge.⁴ In general, constitutive arguments attempt to justify normative requirements by pointing out that we already accept them insofar as we are believers or agents, and so the question "why should I care?" is answered "because you are already committed to following them".

One of the most notable champions of constitutive arguments over the last twenty years is Christine Korsgaard.⁵ The earlier versions of Korsgaard's arguments, especially as presented in *Sources of Normativity*, have received a lot of critical attention.⁶ In response to her critics she has developed a new line of constitutive arguments which figures to help set the agenda on the topic for years to come.⁷ In this paper, I want to discuss these later arguments which haven't yet been discussed as much.⁸ I concentrate on two related but distinct constitutive arguments for the requirement that we must will universally.

Korsgaard's Kantian project is to ground not just rational requirements but also morality in constitutive requirements of agency.⁹ I initially look in some detail at Korsgaard's attempt to arrive at the requirement that we will universally from observations about the causality of the will; the main idea is that one cannot be the cause of one's actions unless one wills universally. This argument has its basis in Kant's argument from the *Groundwork* which relies on the connection between causality and regularity. After reconstructing Korsgaard's argument in section I and showing why this strategy fails in section II, I move on in section III to consider a second argument which attempts to establish that willing universally is constitutive of having a self. In sections IV and V, I look at the instrumental principle as a way of strengthening Korsgaard's argument. Finally, in section VI, I draw on the preceding discussion of Korsgaard to highlight the central challenge facing constitutive arguments in general.

I

Korsgaard claims that because of our reflective nature we don't simply act on a desire or impulse, we consider whether our desires or impulses are *reasons* to act. And in fact, it is because of our reflective nature that we *need* reasons in order to act. A reason, for Korsgaard, is a desire or impulse which can be reflectively endorsed.

When an impulse – say a desire presents itself to us, we ask whether it could be a reason. We answer that question by seeing whether the maxim of acting on it can be willed as a law by a being with the identity in question.¹⁰

The way we decide which of our desires to endorse and which not to, then, is by appealing to a law or principle: "until the will has a law or principle there is nothing from which it can derive a reason".¹¹ And this is where obligation enters in, we are obligated by these laws according to which we endorse our desires or impulses.

Now, even if this were right, it would establish only that we are obligated insofar as we act for a reason – only then, according to the account, do we commit ourselves or appeal to laws in deciding how to act. If we did not act for a reason, we would not be committed to any law and so we would not be obligated. The way Korsgaard gets around this potential problem is by contending that we must always act for a reason. She claims that human beings are such that they need a reason in order to act. As a result, she concludes, *whenever* we act we commit ourselves to a law which obligates us in every other relevantly similar circumstance.

In defending the view that we must act according to laws, Korsgaard develops a line of thought inspired by Kant's argument at the beginning of the Third Section of the *Groundwork*: causality entails laws, and so if the will is to cause our actions then it must be governed by laws.¹² There is, she thinks, a common element in the ordinary notions of causality and of a reason. The ordinary notion of causality combines the idea of power or of one thing affecting another with the idea of *universality, regularity, or law-like formulation*, while the ordinary notion of a reason combines the idea of normativity or of obligation with the idea of *universality, regularity, or law-like formulation*.¹³ Korsgaard follows Kant in rejecting the Humean claim that regularity is all there is to causality while accepting that regularity is necessary to *recognize* causality:

Without endorsing Hume's more reductive and sceptical conclusions, we can agree with him that we could never *identify* the element of necessitation and therefore distinguish cases of causal connection from cases of mere temporal sequence without regularity. And this is part of the story about why we need regularity or law for the idea of causality.¹⁴

Applying this point to the causality of the will, it follows that if we are to think of ourselves as causes of our actions, we have to think of ourselves as acting in accordance with some law.¹⁵

Her argument is highly suggestive but less than fully transparent. Here is a step by step reconstruction of the reasoning behind it, as it seems to me to go. The will acts by endorsing some desire or impulse – this is to be distinguished from desires or impulses directly causing an action without the agent intervening. Because the will's acting is a causal process, and because causality requires regularity or law, the will must endorse a desire or impulse in accordance with a law. But remember, an endorsed desire or impulse *just is* a reason. And so it follows that the acceptance of a reason must be in accordance with a law.

Now, laws are universal in nature. From this, Korsgaard concludes that we must conceive of our reasons as universal in nature as well. And this means that when you endorse a desire or impulse, you must be committing yourself to endorsing a similar desire or impulse in similar circumstances. Which is just to say that you must will universally.

The argument is imaginative and interesting. We can appreciate it even more by seeing how it differs from Kant's. For Kant, *cause* is a concept of the understanding which is supposed to apply only to the world as we experience it – the phenomenal rather than the noumenal realm. The sort of causal relation Korsgaard is focusing on is a relation between something in the world of experience and the will, which for Kant is outside of the world of experience. Kant notices that there is an apparent problem with the causation of the will on his view as it involves a causal relation between something beyond all possible experience and something within the world of experience. In the *Groundwork* he calls it a "causality of a peculiar kind" and has a lengthy discussion of it in the *Critique of Practical Reason*.¹⁶ Because of this, Kant, unlike Korsgaard, cannot (and arguably does not) straightforwardly rely on claims about regularities in order to reach conclusions about how the will "causes" actions.

Of course, Kant's notion of the noumenal realm is notoriously obscure. At least part of the reason Korsgaard's argument is so interesting is that it promises to shed light on the nature of the relation between the causation of the will and the obligation to act according to universal laws *without* invoking this "causality of a peculiar kind". And so, if it works we would get Kantian conclusions through a broadly Kantian argument without some of the most problematic features of the Kantian metaphysics.

II

Korsgaard's strategy, then, is to argue from a premise about the relation between causation and laws to the conclusion that the will must act in accordance with laws of a certain sort. One way to object to this argument would be by challenging this premise and denying that causal relations must be backed by laws.¹⁷ This will not be the path that I will pursue here. I will not be concerned with whether causality requires laws but rather with whether the laws that causality may require are of the sort Korsgaard needs and thinks she has established.

To re-cite a passage from above, Korsgaard says that we could never *identify* the element of necessitation and therefore distinguish cases of causal connection from cases of mere temporal sequence without regularity. And this is part of the story about why we need regularity or law for the idea of causality.¹⁸

There seems to be something like a shift from epistemology to metaphysics here. While in the first sentence Korsgaard is making a claim as to how we identify causal relations, in the second she is making a claim about the very concept of causality. That regularity is necessary to identify a causal relation does not by itself imply that regularity is somehow part of the concept of causality. A microscope may be necessary to identify germs, but this does not imply that a microscope (or the concept of a microscope) is part of the concept of germs.

There are two ways Korsgaard might try to argue for the following thesis:

(A) I cannot regard myself as the cause of my actions without thinking that my actions conform to regularities or laws.

First, she could appeal to an epistemological consideration like:

(1) I need regularity in order to recognize a certain relation as one of causality.

Second, she could appeal to a metaphysical/conceptual consideration like:

(2) The concept of causality contains the concept of regularity.

It's not completely clear which option Korsgaard means to be embracing, and at times she seems to go back and forth between the two.¹⁹

Using (1) to support (A) is not at all promising. (A) asserts that I need to think that my actions conform to regularities; and recall, the sort of regularity at stake for Korsgaard is acting similarly in similar circumstances. If (1) were true, then in order to recognize myself as the cause of my actions it would seem that I would need to have observed a certain regularity in my actions in the past. What sort of regularity, though? One possibility is that in order to recognize that I am the cause of my lying here and now, I would need to have observed that I have lied in similar circumstances in the past. But this would entail a wildly implausible epistemology of action: I would be able to recognize myself as the cause of my actions only after observing myself acting repeatedly in similar ways, and so the first time I chose to lie I would be incapable of recognizing that it was my will that caused the action.

Another possibility is that the regularity relevant to (1) is a regularity between my decisions and my actions. So for example, if I make a decision to lie now, and do lie now, I do not need to have observed a similar decision in the past to presently think of myself as the cause of my lying. All I need to have observed is a regularity between my deciding to ϕ and my ϕ ing, for any ϕ . This is much more plausible, epistemologically speaking. But this sort of regularity cannot be used to establish that to think of myself as the cause of my actions requires regularity in the sense of acting similarly in similar circumstances. After all, there can be a regularity between my deciding to ϕ and my ϕ ing even if what I decide to do changes every time I am presented with a similar circumstance.

While some of Korsgaard's epistemological language seems to suggest that she is using (1) to support (A), I suspect what she really means to be doing is appealing to (2). If (2) were true then in order to think of myself as causing my actions I would need to think of them as conforming to regularities – this is supposed to be a conceptual or metaphysical truth about causation, not an epistemological point.²⁰ This is much more plausible. On this version of the argument, she is not committed to any claims about our having observed regularities in the past. Rather, what she is committed to is that thinking of ourselves as the cause of our actions requires thinking that our actions conform to regularities. But now we get to the central problem for this whole strategy of linking the causation of the will to the requirement that we will universally. Suppose that we do think of the will as causing some action. Even granting Korsgaard her conceptual connection between causality and regularity, all that follows is that we must think of there being a regularity between *will and action* – that is, a regularity between *decision and action*. But this isn't what Korsgaard needs for the sake of universal willing. What she needs is a regularity between *circumstances and decision (or action)*. Such a regularity may follow the causality of circumstances, but this is not what Korsgaard is focusing on. And for good reason: while

there is an indisputable connection between thinking of oneself as an actor and the causality of the will, the same is not the case for the causality of circumstances.

Imagine a world where there are no regularities linking circumstances to decisions (or actions), but there are regularities linking decisions to actions. Given the proposed connection between causality and regularity, it follows that there can be causality of the will in this world, decisions can cause actions, even though there are no laws linking circumstances to decisions. To flesh it out a bit more, imagine that this is how the will functions. It surveys a number of desires looking for one to endorse. This process is not law governed in any way. Once the will chooses a desire to endorse, it now has a reason to act and, let's say, decides to act on that reason. Acting is a causal process, though, and so there must be a law governing the relation between the will and the action. So, a law of action is involved: if I decide to ϕ , I ϕ . But this law does not make any claims as to how desires are to be endorsed or as to whether the will is committing itself to acting similarly in similar circumstances – remember there are no regularities (and thus laws) between circumstances and decisions. Korsgaard's argument has not ruled out this scenario, and without ruling it out she cannot get the conclusion that you must will universally.²¹

To summarize, Korsgaard's argument is that since the will acts by endorsing a desire or impulse, and since the will's acting is a causal process which must be backed by a regularity or law, the will must endorse a desire or impulse in accordance with a law. From this it is supposed to follow that you must will universally. The problem is, the regularity that Korsgaard needs for the sake of her moral theory – a regularity between circumstances and decisions – is not the sort of regularity required for the causality of the will – a regularity between decisions and actions which need not involve circumstances in any way. More generally, the problem raised for Korsgaard is not unique to her account but will face any attempt to argue for the requirement that we will universally on the basis of the causality of the will. The requirement that we will universally involves a relation between circumstances and the will, while the causality of the will involves a distinct relation between the will and action.

To be fair, it may be that Korsgaard thinks these two relations are connected. Perhaps in conceiving of oneself as acting for a reason, one must think of there being a regularity *both* between one's will and one's action, *and also* between one's circumstances and one's will. Some of Korsgaard's arguments discussed below seem to support this claim. Here, my point is just that this conclusion is not established by her present argument regarding the causality of the will, since *prima facie*, any regularity that links circumstances and decisions would seem to be relevant not to the causality of the will (the will causes actions, not circumstances) but rather to the causality of circumstances (the will is, perhaps, the effect of circumstances).²²

III

Although Korsgaard explicitly says that her conclusions about reasons are *dependent* on her view about causality, at certain points she suggests a mere *parallel* or *analogy* between the two.²³ I find this route more promising. She notes that Hume presented two problems: That we do not encounter the power of causes in experience and that we do not encounter the active self in experience. According to Korsgaard, Kant's reply to the first is that the mind

...imposes the notion of causal law on certain temporal sequences in its attempt to understand the empirically given world as a single systematic whole organized in space and time.²⁴

His reply to the second problem, claims Korsgaard, is similar,

...we impose the form of universal volitional principle on our decisions in our attempts to unify ourselves into agents or characters who persist through time – or rather – as I will explain below – who are committed to making the same decision on some range of possible occasions.²⁵

The two cases are similar, she claims, because in both there is an a priori principle that allows us to unite what otherwise would be separate phenomena. The normative principles of the will are supposed to be what bring “integrity and therefore unity – and therefore, really, existence – to the acting self”.²⁶ Unless when I act I project myself into other similar circumstances and commit myself to acting similarly, I do not have a unified self.

What is this unified self? Korsgaard places a great emphasis on the distinction between me or my will acting and desires or impulses acting in me.²⁷ When she claims that having a unified self requires that when I act I project myself into other similar circumstances and commit myself to acting similarly, she is claiming that without this sort of projection it cannot be me or my will that is acting, it can only be desires or impulses acting in me.

For if *all* of my decisions were particular and anomalous, there would be no identifiable difference between *my acting* and *an assortment of first-order impulses being causally effective in or through my body*. And then there would be no self – no mind – no me – who is the one who does the act.²⁸

Again, Korsgaard's argument which is supposed to lead us from the assumption of reflective distance to a requirement of willing universally is less than fully transparent. Here is what seems to me to be the strongest way of developing it. Unless I will universally I don't really have a self at all. But if I have no self there is nothing to stand in reflective distance over my desires. We already know that I *do* have reflective distance over my desires (this is the assumption Korsgaard starts out with). Therefore, I must will universally.²⁹

One might question Korsgaard's distinction between me acting and desires or impulses acting in me. It seems at least somewhat plausible that insofar as it is *my* desires that are causally efficacious, it is *me* acting. However, I will not press this point here. One also might question whether we can really have reflective distance from our desires, that is, whether we can reflect on our desires and choose which ones to act on. But this is Korsgaard's starting point – again, she thinks that the source (and the solution) of the normative problem is our reflective nature – so I think that it is more fruitful to grant the point. What I am going to question is the move from our reflective distance from our desires to a requirement that we will universally or that we appeal to a law in order to endorse our desires. A type of reflective distance from my desires seems possible even if there is no universality or regularity of the sort described by Korsgaard. All that seems to be needed for reflective distance is an ability to endorse or reject our desires, whether according to laws or not. In effect, I will be denying the first step of the argument presented above. The crucial issue is this: If I can reject even my strongest desire, then there is a difference between my self acting and desires acting in me – this alone could ground a type of

reflective distance and thus provide a self that is distinct from desires or impulses, even by Korsgaard's own standards. But the ability to reject even my strongest desires does not seem to require that I project my action into other similar circumstances – that is, contrary to Korsgaard, it does not seem to require that I will universally. Or so I will argue below.

IV

It is possible to develop Korsgaard's argument and perhaps even preempt the objection just hinted at by looking at some of Korsgaard's remarks concerning the normativity of the instrumental principle.³⁰ Korsgaard thinks that the will gets its unity from normative principles. The universality requirement discussed above is one such principle; the requirement that one take the means to one's ends is another. To support her claim that the obligation of willing universally derives in some sense from the unity of the self, she discusses what she thinks may be a less controversial case, that of the instrumental principle.

Korsgaard argues that the normativity of the instrumental principle can only be explained by appealing to a commitment that is constitutive of the self. That is, unless you are committed to following the instrumental principle, there is no you who is acting. Suppose that I will some end. Korsgaard claims that if I were to change my mind about whether to pursue that end every time I had a new desire or whenever the means to fulfilling that end got too hard, then I wouldn't really be willing that end in the first place. If I always change my mind according to the desire of the moment then there is no distinction between *me* acting and desires and impulses acting in me. And so, Korsgaard thinks that a commitment to following the instrumental principle is constitutive of willing.³¹

Again, Korsgaard discusses the instrumental principle to provide an example of how a normative principle could be constitutive of the will. This is meant to serve as an analogy for how the principle that we must will universally could also be constitutive of the will. But ultimately she relies on her discussion of the instrumental principle as more than just an analogy; she argues that the instrumental principle cannot stand alone without a principle for determining one's ends. I will eventually return to this point. For now though I want to focus on Korsgaard's discussion of the instrumental principle in order to clarify her use of it.

The instrumental principle is normative, Korsgaard assumes. She argues that its normativity cannot be explained on either an empiricist model, like Hume's, or a dogmatic rationalist model, like Clarke's. I will not be concerned here with her arguments against these two views. I will assume that she is right in claiming that the instrumental principle is a normative principle and that the normativity of the instrumental principle can only (or can best) be explained on a Kantian view like the one she proposes.

For Korsgaard the instrumental principle is normative because in willing an end we commit ourselves to pursuing the means to that end; that is what willing an end *is*. And so when we fail to pursue the means we are violating that commitment which we ourselves made. One might think that in failing to pursue the means we are also thereby dropping the end. But Korsgaard argues that if failing to pursue the means implied that we are giving up the end, the instrumental principle would not be normative because we could not violate it – it would never be the case that one fails to pursue the means to one's ends.³² She thinks that we can get an independent fix on what someone's ends are in other ways – that is, aside from seeing what means she pursues – for example, by asking her. So for Korsgaard, the

normativity of the instrumental principle depends on one's commitment to pursuing the means insofar as one wills a certain end.

Notice that there are a couple of ambiguities here. The first is an ambiguity about how "pursuing" should be understood. The instrumental principle could be understood either as the requirement that you *will* the means to your ends or the requirement that you *take* the means to your ends.³³ The second is an ambiguity about what is constitutive of willing. It either could be that it is *commitment to* (or perhaps *acceptance of*) the instrumental principle that is constitutive of willing, or it could be that it is *conformity* to the instrumental principle that is constitutive of willing.³⁴

If *conformity* to the instrumental principle were constitutive of willing, *you* could not violate the instrumental principle. Insofar as "you" violated the principle, there would be no you. Or rather, if conformity to the instrumental principle so understood were constitutive of willing, one could not *will* in a way that violated the instrumental principle; insofar as one *willed* at all, one would be conforming to it. There would be no room for someone who fails to take (or will) a given means but nevertheless wills the corresponding end. But if so, then Korsgaard's view would fail to account for the normativity of the instrumental principle: the instrumental principle would not be a principle that one can fail to conform to – to fail to take (or will) the means is to fail to will the corresponding end, but if you do not will that end then the instrumental principle does not apply to you with respect to that end. This is precisely the sort of objection she raises with respect to the empiricist view – it cannot account for the normativity of the instrumental principle because it does not allow for the possibility of violating the principle.³⁵

Her view must be that it is a *commitment to* or *acceptance of* the instrumental principle which is constitutive of willing. So the requirement is essentially a requirement that one will (or commit oneself to) taking the means one believes are necessary for the ends one wills or intends to pursue.³⁶ And one is required to intend to take the means to one's ends because one is committed to doing so.

I now want to tie this all back to the point I started this section with. Again, Korsgaard thinks that the will gets unity (and thus existence) from normative principles like the instrumental principle. Willing the means to one's ends is constitutive of the very act of willing, and so, a self that is not obligated by the instrumental principle is not a self at all. This is the point Korsgaard is making (though she puts things in terms of the hypothetical imperative instead of the instrumental principle) when she writes,

The reason I must conform to the hypothetical imperative is that if I don't conform to it, if I *always* allow myself to be derailed by timidity, idleness, or depression, then I never really *will* an end. The *desire* to pursue the end and the desires that draw me away from it each hold sway in their turn, but *my will* is never active. The distinction between my will and the operation of the desires and impulses in me does not exist, and that means that I, considered as an agent, do not exist. Conformity to the hypothetical imperative is thus constitutive of having a will. It is, in fact, an essential part of what gives you a will.³⁷

Although there are some possible problems with this account of the normativity of the instrumental principle that I have not discussed here, one can see how this view may seem attractive.

However, notice that one can accept that the instrumental principle is constitutive of willing, and thus that it allows for a distinction between my acting and desires acting in me, without accepting the requirement that we must will universally.³⁸ Why not appeal to Korsgaard's view about the instrumental principle to support the objection I suggested at the end of the previous section? The objection is that one may be able to draw Korsgaard's distinction between my self acting and desires acting in me without appealing – as she does – to the requirement that I will universally. Given that when I will an end I am also committed to taking the means to that end, according to Korsgaard's own view, my willing that end is distinguished from a desire for that end acting in me and willing universally need not be part of the picture.

In other words, Korsgaard's discussion of the instrumental principle seems to undermine her point about the principle of willing universally, not support it. If the instrumental principle allows for the distinction between me acting and desires acting in me, as Korsgaard claims, then the principle of willing universally does not seem to be required for having a self.

V

Korsgaard has a response available to this sort of objection. She argues that the instrumental principle cannot stand alone, but requires some further principle for determining ends.

... the normative force of the instrumental principle does seem to depend on our having a way to say to ourselves of some ends that there are reasons for them, that they are good.³⁹

Her idea seems to be that if one can will any end at any time, drop one's ends or change them without a reason, then the requirement to take the means to one's ends would be somewhat vacuous.

If I am to will an end, to be and to remain committed to it even in the face of desires that would distract and weaknesses that would dissuade me, it looks as if I must have something to *say to myself* about why I am doing that – something better, moreover, than the fact that this is what I wanted yesterday. It looks as if the end is one that has to be *good*, in some sense that goes beyond the locally desirable. I have to be able to make sense to myself of effort and deprivation and frustration, and it is hard to see how the reflection that this *is* what I wanted yesterday can do that by itself, especially when I want something else today.⁴⁰

Note however that the instrumental principle only requires that one be committed to the means to the ends one intends to pursue. There are two ways to fulfill this requirement: either you become committed to the means or you give up the end. There is no requirement to take the means to one's ends unconditionally; one can change one's mind. And there need not be anything wrong with that. The instrumental principle does not require you to maintain your ends over a period of time, it only requires you to keep some *consistency* between your current intentions.⁴¹ So certainly Korsgaard is not arguing that the problem with not having a principle to determine one's ends is that without such a principle one would be violating the instrumental principle. What her argument needs to be is that in the absence of a principle to determine one's ends – that is, the requirement of willing universally – the instrumental principle alone would not be enough to provide unity and therefore existence to the self.

Korsgaard seems to be assuming that if you don't will universally then you would always change your mind whenever a new desire arises – and if no new desire arises, well, that is just an accident.⁴² In that case, there again would be no distinction between you acting and desires acting in you. The first thing to say in response to this is that the point I made at the end of section III seems still to apply, perhaps even more forcefully than before. A type of reflective distance seems possible even if one does not will universally: desires could be endorsed or rejected without doing so according to laws. In fact, armed by the preceding discussion of the instrumental principle, it seems we can draw a distinction between willing an end and a desire for that end acting in me in the following way. When I will an end I take on a commitment to take the means to that end; but when I have a desire there is no commitment I need to take on. But why not try to use *this* sort of commitment – as opposed to the commitment involved in willing universally – to ground the distinction between my acting and desires acting in me, regardless of whether I'm changing my mind all the time, or as a matter of mere accident not doing so?

Imagine Jim, a student who adopts ends without any reason and often changes his mind as to what ends to adopt. His first semester, he decides that he wants to become a doctor; his second semester a writer; his third semester an actor; and so on. Jim takes his current ends very seriously, diligently pursuing the means to the ends he currently has. For instance, after his decision to become a doctor, he takes all the required pre-med classes and starts to study for the MCATs. Now, when adopting an end, Jim does not commit himself to maintaining the end until he sees good reason to change his mind – that is, his adoption of ends is not based on reasons and he does not will universally. Still, it does not seem that Jim's adopting of ends needs to be produced by new desires in him directly causing him to act. It seems at least possible that each time Jim adopts an end, he takes on commitments to take the means required for achieving his goals. While his commitments may sometimes be short lived they are commitments nonetheless, which is enough to guarantee by Korsgaard's own standards that it is *he* who is acting as opposed to desires acting in him.⁴³ This may not be the most desirable way to be, one may not accomplish much. But surely the problem with this person is not that he has no self – again, even on Korsgaard's own view of what having a self amounts to.

In discussing Korsgaard's view, Bratman argues that there are two possible interpretations of what Korsgaard is up to: the first focuses on the unity of agency at the time of the action, while the second focuses on the relationship between the current decision and future ones.⁴⁴ Indeed, throughout Korsgaard's earlier works one can't avoid a suspicion that she is moving back and forth between the question of what constitutes a self at a particular moment and the issue of personal identity over time. In her latest work, however, Korsgaard argues that the two issues are related, she makes this point in discussing the example of the socialist Russian nobleman given by Parfit.⁴⁵ The young Russian nobleman plans to distribute among the peasants large portions of the inheritance he will be receiving years later. But he also anticipates that he will change his mind about this as he becomes older and his values more conservative. Korsgaard claims that the Russian nobleman provides an example of "particularistic willing" – a failure of willing universally – because he adopts an end (to distribute the money to the peasants) and yet expects to change his mind without a reason. She says:

He can decide to disagree with his own future attitude. But *unless* he is then also prepared to regard his own future attitude as one of weakness or irrationality, he is not according the reason he himself proposes to act on *right now* as having normative standing. For he is not making a law for himself unless he thinks of his future attitude as a violation of that law, and if he does not think he can make laws for himself then he lacks self-respect. So his problem is not his disunity with his future self, but his disunity with himself here and now.⁴⁶

The example of Jim given here parallels the Russian nobleman. I take Jim to be committing himself to taking the means to achieving his career goals without thereby thinking that he is committed to these goals until he finds a reason to change his mind. Rather, his commitment is to pursuing the means for as long as he maintains his end. I take it that Korsgaard would say about Jim what she says about the Russian nobleman: insofar as he does not will universally, Jim is showing disunity with himself at that moment, not simply with his future self. My point here is that he may well be unified at that moment in wholeheartedly pursuing his current end.

Perhaps one may argue that Jim has *less* of a unified self than someone who wills universally.⁴⁷ Korsgaard moves in this direction of discussing degrees of selfhood in *The Constitution of Agency*:

People are more or less successful at constituting their identities as unified agents, and a good action is one that does this well... But since action requires agency, it follows that an action that is less successful at constituting its agent is to that extent less of an action. So on this conception, "action" is an idea that admits of degrees. An action chosen in a way that more successfully unifies and integrates its agent is more authentically, more fully, an action, than one that does not. And this in turn is where the principles of practical reason, the hypothetical and categorical imperatives, come in to the story.⁴⁸

The problem with this proposal from Korsgaard's standpoint is that it would seem to require her to significantly rework her project of grounding normativity. Whereas the force of Korsgaard's original constitutive argument derives in large part from the plausibility of her premise that everyone has a self, a reworked argument dealing with *degrees* of selfhood would instead need to rely on the premise that everyone has a *robust* self. This is problematic. Remember, Korsgaard's original argument depends on the claim that unless I have a self, there is nothing to stand in reflective distance over my desires. The reworked argument would thus need to claim that unless I have a *robust* self, there is nothing to stand in such reflective distance. But why couldn't a non-robust self do the job? If my discussion of the instrumental principle is correct then there is a way of making a distinction between me acting and desires acting in me *even if* I have a non-robust self which fails to will universally – the instrumental principle by itself seems to provide my non-robust self with the reflective distance sought.

Maybe it is more desirable to have a robust self than a non-robust self and maybe having a robust self depends on willing universally. Notice, though, that this is a normative claim and so it cannot be used to *ground* normativity, meaning that it is of no use to Korsgaard in the present context. Moreover, the revised argument would lose the appeal of the most influential constitutive arguments, which seek to derive normative conclusions from uncontroversial non-normative premises.

Alternatively, perhaps Korsgaard's view is that to have a self requires merely that one be *disposed* towards unity. In that case, the extent to which one actually achieves unity may vary, and perhaps one could get by with a fairly minimal self – or at least Korsgaard could grant this. And then insofar as one has a self, one must be disposed to become more unified. If so, and if, as I granted above, willing universally makes one more unified, then one who wills universally is more fully realizing this disposition towards unity which is constitutive of selfhood. And perhaps this is enough to ground the normativity of the requirement of willing universally.⁴⁹

What would be the exact nature of the disposition in question though? Does one *always* need to will universally in order to qualify as having the disposition towards unity? If so, then the proposal is equivalent to saying that one must always will universally in order to have a self at all, and as I argued with the Jim example, this seems implausible. Alternatively, maybe all that is needed to qualify as having the disposition is that one *sometimes* or *often* will universally. But in that case, the proposal would not seem to support categorical moral requirements. For suppose one already wills universally often enough that one already gets credit for the disposition toward unity, but now is confronted with a new decision. Then one could choose not to will universally in making this new decision without losing credit for the disposition – one still has the general disposition, it just is not expressed in this particular case – and so without compromising one's selfhood. But then, the need to have a self would not provide support for the requirement to will universally in this case.⁵⁰

If the objections I have set out against Korsgaard are successful, where does that leave us? Recall, Korsgaard thinks that we cannot act unless we act for a reason. And acting for a reason means reflectively endorsing a certain desire in accordance with a law which we give to ourselves. So, whenever we act we commit ourselves to a law requiring us to act similarly in similar circumstances. There are alternatives to this picture though. For instance, one may hold with Korsgaard that having a reason requires reflective endorsement based on a law, but disagree with her by denying that we must always act for a reason. Alternatively, one may hold with Korsgaard that having a reason requires reflective endorsement, but disagree with her by denying that such endorsement must be based on a law. Finally, one may disagree with Korsgaard on both counts.⁵¹ In effect, what I hope to have shown is that Korsgaard has not given us compelling reasons to accept her picture of things rather than these others.

VI

We have seen a similar dynamic arise in the discussion of each of the two constitutive arguments from Korsgaard. The challenge for making a successful constitutive argument is to start with a modest premise that is nearly universally accepted, and manage to derive from it a substantial conclusion by relying on constitutive claims. But this is really hard to do. Often, a modest premise is capable of yielding only a modest and uninteresting conclusion. So for instance, that we are the causes of our actions or that we have selves at all are modest enough claims to be relatively uncontroversial starting points. The trouble with them is that nothing of real ethical substance seems to follow from them together with the relevant constitutive claims. In the case of our being the cause of our actions, one may be able to derive from this thesis the conclusion that there must be some sort of regularity between

deciding and acting, but not the sort of ethically relevant regularity that Korsgaard wants to get. In the case of having a self, one may be able to derive from this thesis the conclusion that one must accept some principle of practical reason, but not necessarily the principle of willing universally.

To try to avoid this sort of pitfall, a philosopher making a constitutive argument may be tempted to begin with a more substantial starting point, capable of yielding a more substantial conclusion. For instance, in the case of Korsgaard's argument regarding our being the causes of our actions, one may try to start with the claim that the regularity in question holds between circumstances and decisions (or actions). While in the case of her argument regarding having a self, one may try to start with the claim that we have a fairly robust self. Now the problem is that too much is being built into the front end of the argument: its initial assumptions ends up being at least as controversial as the conclusion they are meant to support.

The danger, in making constitutive arguments, is to start with a substantial point that sounds more modest than it really is. This seems to be especially easy to do when one is dealing with certain philosophically central but disputed terms like "action" or "self". For instance, one might start with the innocuous claim that we have a self. Few would deny this. What's more, if through a kind of alchemy one could somehow manage to derive moral requirements from this platitudinous claim, *that* would be genuinely exciting. This is precisely the sort of possibility that generates enthusiasm for constitutive arguments in the first place. But then problems arise: deriving such requirements turns out to be like squeezing water from a stone and something more substantive gets built into the assumptions. So the assumption that we have a self becomes the assumption that we have a *robust* self. This new assumption may well yield moral requirements, but only by being as controversial as those requirements. Having a robust self might be *desirable* (even if we don't all have it), but one cannot rely on this normative point in attempting to ground moral normativity. The successful constitutive argument is one that navigates between these two pitfalls as opposed to surreptitiously oscillating between them.⁵²

References

- Anscombe, G.E.M.. 1981. "Causality and Determination." *The Collected Papers of G.E.M. Anscombe, Volume Two: Metaphysics and the Philosophy of Mind* (Minneapolis: University of Minnesota Press)
- Bratman, Michael E.. 1998. "The Sources of Normativity." *Philosophy and Phenomenological Research* 58, No. 3: 699-709. Reprinted in Michael Bratman, *Faces of Intention* (Cambridge: Cambridge University Press, 1999)
- Broome, John. 2000. "Normative Requirements", in Dancy ed., *Normativity*, Blackwell.
- Cullity, Garrett and Gaut, Berys eds.. 1997. *Ethics and Practical Reason*. Oxford: Clarendon.
- Dreier, James. 1997. "Humean Doubts about the Practical Justification of Morality." In Cullity and Gaut eds..
- 2001. "Humean Doubts about Categorical Imperatives", in Elijah Millgram ed., *Varieties of Practical Reason* (Cambridge, MA: MIT Press).
- Enoch, David. 2006. "Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Action", *Philosophical Review* 115, No. 2.
- Ferrero, Luca. 2009. "Constitutivism and the Inescapability of Agency." Russ Schafer-Landau ed., *Oxford Studies in Metaethics*, vol. 4. (Oxford: Oxford University Press).
- Hill, Thomas E. Jr.. 1973. "The Hypothetical Imperative." *Philosophical Review* 82, No. 4.
- Kant, Immanuel. 1993. *Critique of Practical Reason*. Translated by Beck. New Jersey: Prentice Hall. (References are to the Academy edition volume, page number)
- 1997. *Foundations of the Metaphysics of Morals*. Translated by Beck. New Jersey: Prentice Hall. (Referred to as *Groundwork* followed by the Academy edition volume, page number)
- Korsgaard, Christine M.. 1989. "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs* 18, No. 2: 101-132. Reprinted in Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996).
- 1996. *The Sources of Normativity*. Cambridge, MA: Cambridge University Press.
- 1997. "The Normativity of Instrumental Reason." In Cullity and Gaut. Reprinted Korsgaard (2008).
- 1999. "Self-Constitution in the Ethics of Plato and Kant." *The Journal of Ethics* 3: 1-29. Reprinted in Korsgaard (2008).
- 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.
- 2009. *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University Press, 2009).
- Lavin, Douglas. 2004. "Practical Reason and the Possibility of Error." *Ethics* 114: 424-457.
- O'Hagan, Emer. 2004. "Practical Identity and the Constitution of Agency." *The Journal of Value Inquiry* 38: 49-59.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Railton, Peter. 1997. "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action." In Cullity and Gaut. Reprinted in Peter Railton, *Facts*,

Values and Norms: Essays Towards a Morality of Consequence (Cambridge: Cambridge University Press, 2003)

- Rosati, Connie S.. 2003. "Agency and the Open Question Argument." *Ethics* 113: 490-527.
- Setiya, Kieran. 2003. "Explaining Action." *Philosophical Review* 112, no. 3: 339-393.
- Skidmore, James. 2001. "Skepticism about Practical Reason: Transcendental Arguments and Their Limits." *Philosophical Studies* 109: 121-141.
- Sobel, David. 2001. "Subjective Accounts of Reasons for Action." *Ethics* 111: 461-492.
- Velleman, J. David. 1992. "What Happens When Someone Acts?" *Mind* 101: 461-81. Reprinted in Velleman (2000).
- 1996. "The Possibility of Practical Reason." *Ethics* 106: 694-726. Reprinted in Velleman (2000).
- 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press
- 2009. *How We Get Along*. Cambridge: Cambridge University Press.
- Wallace, R. Jay. 2001. "Normativity, Commitment, and Instrumental Reason." *Philosophers' Imprint* 1, No. 3
- 2004. "Constructing Normativity." *Philosophical Topics* 32; Nos. 1 and 2: 451-476.

¹ See for instance: Velleman (1992), (1996), (2000), (2009); Korsgaard (1996), (1997), (2008), (2009); Railton (1997); Dreier (1997), (2001); Skidmore (2001); Rosati (2003); Setiya (2003); O'Hagan (2004); Wallace (2004); Enoch (2006); Ferrero (2009).

² For instance, Korsgaard (1996), (2008) argues that commitments to following certain principles are constitutive of action and belief; Velleman (2000), (2009) argues that certain aims are constitutive of action and belief; Railton (1997) discusses both types of argument.

³ See for example: Korsgaard (1996), Velleman (2000), Rosati (2003).

⁴ See for example: Korsgaard (1996), Velleman (2000), Railton (1997).

⁵ Korsgaard (1989), (1996), (1997), (1999), (2008), and (2009).

⁶ See for instance the replies by Cohen, Geuss, Nagel, and Williams contained in Korsgaard (1996). See also Bratman (1998) and Wallace (2004).

⁷ This new line is first advanced in reply to Nagel and Cohen in Korsgaard (1996) and further developed in Korsgaard (1997), (1999), (2008), and (2009).

⁸ But see for example Enoch (2006).

⁹ Unlike Kant, however, Korsgaard does not take the requirement of willing universally to be sufficient to establish morality, she thinks that justifying this requirement is the first step towards justifying morality. See for example Korsgaard (1996: § 3.2.4).

¹⁰ Korsgaard (1996: 113). See also (1996: 97): "'Reason' means reflective success. So if I decided that my desire is a reason to act, I must decide than on reflection I endorse that desire."

¹¹ Korsgaard (1996: 98)

¹² "Since the concept of a causality entails that of laws according to which something (i.e., the effect) must be established through something else which we call cause, it follows that freedom is by no means lawless even though it is not a property of the will according to laws of nature. Rather, it must be a causality of a peculiar kind according to immutable laws." (*Groundwork* IV, 446)

¹³ Korsgaard (1996: 225-6)

¹⁴ Korsgaard (1996: 227)

¹⁵ "For to regard oneself as an agent is to regard oneself as a cause, as productive of certain actions and their effects. And given the connection between causality and regularity, to do that must be to regard oneself as productive of these actions and effects in some regular way. This is what Kant means when he says that since the will is a cause it must operate according to a law." (Korsgaard (1996: 227))

¹⁶ See the passage from *Groundwork* IV, 446 cited above and see for example, *Critique of Practical Reason* V, 49-57.

¹⁷ For a discussion of the possibility of causation without laws, see Anscombe (1981).

¹⁸ Korsgaard (1996: 227)

¹⁹ See her discussion in Korsgaard (1996: 225-7).

²⁰ Korsgaard seems to support the conceptual claim with epistemological considerations: in short, she uses (1) to support (2). See for example the passage from Korsgaard (1996: 227) cited above. If so, then the problems raised above for (1) threaten to reoccur for (2). But I am setting this aside to discuss (2).

²¹ In effect, this may be part of what Nagel and Cohen are getting at in their replies to Korsgaard's *Sources of Normativity* Tanner lectures. Nagel questions the necessity of deciding according to laws: Why can't the self-determining will, determine itself in "individual, disconnected choices as well as according to some consistent law?" (Korsgaard (1996: 202)) And Cohen asks why reflective endorsement must involve laws: "The reflective structure of human consciousness may require ... that, on pain of reducing myself to the condition of a wanton, I endorse the first-order impulses on which I act.... But it does not follow, and it is not true, that the structure of my consciousness requires that I identify myself with some law or principle ... sometimes the commands that I issue will be singular, not universal." (Korsgaard (1996: 176)) The argument of Korsgaard that I have been presenting is meant to be a response to Nagel and Cohen. In effect part of what I am claiming is that it is an unsuccessful response.

²² I thank a reviewer for *Philosophical Quarterly* for calling to my attention this possibility.

²³ For her dependence claim see for instance Korsgaard (1996: 227). For her claims about a parallel see for instance Korsgaard (1996: 228-9).

²⁴ Korsgaard (1996: 229)

²⁵ Korsgaard (1996: 229)

²⁶ Korsgaard (1996: 229). See also Korsgaard (2009: 25): "The necessity of conforming to the principles of practical reason comes down to the necessity of being a unified agent."

²⁷ See for instance Korsgaard (1996: 227): "To will is not just to be a cause, or even to allow impulse in me to operate as a cause, but, so to speak, to consciously pick up the reins, and make myself the cause of what I do. And if I am to constitute myself as the cause of an action, then I must be able to distinguish between my causing the action and some desire or impulse that is 'in me' causing my body to act. I must be able to see myself as something that is distinct from any particular, first order, impulses and motives, as the reflective standpoint in any case requires. Minimally, then, I am not the mere location of a causally effective desire but rather am the agent who acts on the desire. It is because of this that if I endorse acting a certain way now, I must at the same time endorse acting the same way on every relevantly similar occasion."

²⁸ Korsgaard (1996: 228). In later works, she refers to failing to will universally as "particularistic willing" and argues against it on similar grounds. See for example Korsgaard (1998: 27) and Korsgaard (2009: §4.4).

²⁹ As a point of clarification, one may wonder whether Korsgaard's argument is trying to establish the claim that I must *sometimes* will universally. This may be suggested by the quotation above from p. 228. However, if this were so, then Korsgaard's argument would only yield the requirement that one sometimes will universally, and this is unlikely to yield the categorical moral requirements

that she is ultimately trying to establish. Her argument is that unless I will universally at a particular moment, it is not *me* who is acting (this is clear, for example, in the passage from p. 227 cited in footnote 27). And so the claim that she is trying to establish with this argument is that I must *always* will universally.

³⁰ Korsgaard gestures at this sort of move in Korsgaard (1996: 230). I am here developing her argument by appealing to Korsgaard (1997) and (2009). She also appeals to the same line of thought in Korsgaard (1999).

³¹ See for example, “The final answer, then, to the question – what gives the instrumental principle its normativity? – is this: conformity to the instrumental principle is an essential part of what makes you a person. There is no position from which you can reject the government of instrumental reason: for if you reject it, there is no you.” (Korsgaard 1997: 254)

³² This discussion assumes that for a principle of practical reason to be normative it must be possible to violate it. For a discussion of this assumption, as well as Korsgaard’s defense of it, see Lavin (2004).

³³ Kant seems to have held the first view, not the second; for a discussion of Kant’s view see for instance Hill (1973). Korsgaard discusses Kant’s view in Korsgaard (1997: §3).

³⁴ Korsgaard varies between these formulations of the instrumental principle, see Korsgaard (1997: 244-254).

³⁵ In arguing against Hume’s view, she says, “If we don’t make a distinction between a person’s ends and what he actually pursues, it will be impossible to find a case in which he violates the instrumental principle” (Korsgaard 1997: 230)). For a critique of Korsgaard’s claim that Hume cannot allow for a distinction between a person’s ends and what she actually pursues see Sobel (2001).

³⁶ For simplicity of exposition, I often drop the “that one believes are necessary” out of the formulation of the instrumental principle. I am assuming that the instrumental principle requires one to take what one believes are the means to one’s ends (even if one’s beliefs are mistaken). This is because on the views under discussion the instrumental principle is a requirement on a combination of mental attitudes, not a requirement that one know how to best accomplish one’s ends.

³⁷ Korsgaard (2009: 69)

³⁸ Dreier (1997), (2001) argues for a view that is in some ways similar. He argues that a certain version of the means-ends principle (not quite Korsgaard’s) is constitutive of agency but resists the claim that moral principles are constitutive of agency.

³⁹ Korsgaard (1997: 251)

⁴⁰ Korsgaard (1997: 250)

⁴¹ It’s not that Korsgaard disagrees with this, “The hypothetical character of the principle implies that you can actually conform to it in either of two ways: you may take the means, or you may cease to will the end” (Korsgaard (1997: 237)). For an especially clear discussion of this point, see Broome (2000).

⁴² See for example Korsgaard (2009: 169). After describing Jeremy, who changes his mind about what to do every second, she writes, “Of course the democratic life does not have to be like this; it is only an accident that each of Jeremy’s impulses leads him to an action that completely undercuts the satisfaction of the last one. But that is the trouble, for it is also only an accident if that does *not* happen. The democratic person has no resources for shaping his will to prevent this, and so he is at the mercy of accident. Like Jeremy, he may be almost completely *incapable of effective action*”.

⁴³ Remember that Korsgaard thinks we can get an independent fix on what someone’s ends are in other ways than by looking at whether she pursues the means to that end, for example by asking her.

⁴⁴ Bratman (1998: 706). His discussion pertains Korsgaard (1996), (1997).

⁴⁵ Korsgaard (2009: §9.7.2), Parfit (1984: 327).

⁴⁶ Korsgaard (2009: 203)

⁴⁷ I want to thank a reviewer for The British Society for Ethical Theory for pressing me on this point.

⁴⁸ Korsgaard (2009: §1.4.8)

⁴⁹ I want to thank a reviewer for the *Philosophical Quarterly* for suggesting this possibility to me.

⁵⁰ For a related point, see note 29 above.

⁵¹ Notice that the point that Korsgaard is defending and that I have been criticizing here is not simply that reasons are general or that “thought traffics in general”. She says that “[t]his by itself... does not quite commit me to reaching the same conclusion about what it would be appropriate to do on all relevantly similar occasions as I reach now.” (1996: 228) Thanks to a reviewer for The British Society for Ethical Theory for prompting me to be clearer on this.

⁵² I want to thank the participants of the Research Triangle Ethics Circle in Chapel Hill and those of the Conference of the British Society for Ethical Theory in Southampton for helpful discussions of earlier versions of this paper.