THE VALUE OF HUMANITY: REFLECTIONS ON KORSGAARD'S TRANSCENDENTAL ARGUMENT

ROBERT STERN

The purpose of this paper is not to consider the worth of transcendental arguments in general (which I have done at length elsewhere),[1] but instead to focus on a specific example of the genre. However, this is an example taken not from epistemology or metaphysics where (in Anglo-American philosophy at least)[2] such arguments have most usually found a home, but rather from ethics. Nonetheless, I hope that what I have chosen to do will not prove valueless, as although specific and in some ways untypical, the argument I set out to discuss is important and influential, but not much considered as a transcendental argument as such. The argument in question is Christine Korsgaard's, from her book *The Sources of Normativity*. My aim is to try to understand what role her transcendental argument for the value of humanity is meant to play in her project, and whether the argument succeeds. Rather to my surprise, and rather against the run of the critical literature on Korsgaard's book, I will suggest that in one of its forms, the

1 Most particularly in (Stern 2000).

2 Transcendental arguments have of course been used by Habermas, Apel and others as part of their projects in social philosophy. For a discussion of current uses of transcendental arguments in ethics, but which curiously hardly mentions Korsgaard, see (Illies 2003).

argument can be made to work—at least in its own terms[3] and when its rather limited place in Korsgaard's overall strategy is understood. And in the end, of course, it may also turn out that there are general lessons to be gained from the examination of this argument after all.

I

Taken as a whole, Korsgaard's aim is to show that we stand under moral obligations, by constructing an argument to that effect. But only part of the overall argument is meant to be a transcendental one—roughly, the middle part. How that specifically transcendental argument is understood therefore depends on how one conceives of what precedes it, and what work its conclusion is supposed to do in what follows. Let me begin by setting out in broad terms what I take the three phases of Korsgaard's overall argument to be.

Phase One: From free agency to the categorical imperative

Korsgaard starts from a Kantian antinomy: that we conceive of ourselves as free on the one hand, but also as agents with a certain stability of purpose and character on the other, which means we take our actions to be governed by principles or laws. Korsgaard's Kantian solution to this antinomy is to argue that these principles or laws must be self-imposed.[4] But now we seem

---

3 For example, as we shall see further below, I think the argument relies heavily on other arguments Korsgaard gives against realism, where in this discussion I allow her to take these for granted (but where I am critical of them elsewhere).

4 Cf. (Korsgaard 1996b: 97-8): '[Kant] defines a free will as a rational causality which is effective without being determined by an alien cause, including the desires and inclinations of the person. The free will must be entirely self-determining. Yet, because the will is a causality, it must act according to some law or other. Kant says: 'Since the concept of a causality entails that of laws…it follows that

to face a second antinomy: on the one hand, unless a principle or law already binds the will, on what basis can the will rationally legislate a law to itself; but on the other hand, if the will is already bound by a law, how can this count as self-legislation?[5] Korsgaard's proposal is that Kant's notion of the categorical imperative is designed to resolve this second antinomy: on the one hand, the free will is not completely lawless, because it must act in accordance with a law or principle; but on the other hand, this does not constrain it or make it less free, because it is just constitutive of free legislation that it has this structure,[6] while which law it chooses is left open.[7]

---

freedom is by no means lawless…' Alternatively, we may say that since the will is practical reason, it cannot be conceived as acting and choosing for no reason. Since reasons are derived from principles, the free will must have a principle. But because the will is free, no law or principle can be imposed on it from the outside. Kant concludes that the will must be autonomous: that is, it must have its *own* law or principle.'

5 Cf. (Korsgaard 1996b: 98): 'And here again we arrive at the problem. For where is this law to come from? If it is imposed on the will from outside then the will is not free. So the will must make the law for itself. But until the will has a law or principle, there is nothing from which it can derive a reason. So how can it have any reason for making one law rather than another?'

6 Cf. (Korsgaard 1996b: 235): 'Cohen makes it sound as if autonomous lawmaking were one thing, and universal autonomous lawmaking another, and this in turn makes it sound as if universalizability is a rational constraint which is imposed on what would otherwise be the arbitrary unconstrained activity of autonomous lawmaking. But I think Kant himself means something else, namely autonomous lawmaking just *isn't* autonomous lawmaking unless it is done universally. The requirement of universalization is not imposed on the activity of autonomous lawmaking by reason from outside, but is constitutive of the activity itself'.

7 Cf. (Korsgaard 1996b: 98): 'The problem faced by the free will is this: the will must have a law, but because the will is free, it must be its own law. And nothing determines what that law must be. *All that it has to be is a law*. Now consider the content of the categorical imperative, as represented by the

The conclusion of this first phase of the argument is therefore that in order for the will to be free, it must act on the basis of something that has the nature of a law or principle.

Before moving on to Phase Two of Korsgaard's overall argument, it may be worth pausing to emphasise that (at least as I see it) Phase One is not a transcendental argument: rather, it is an argument that works by showing how the antinomies of free agency and of self-legislation need to be resolved, leading to the categorical imperative, to act only on a maxim that I can will as a universal law.

Phase Two: From the categorical imperative to the moral law, step 1: the value of your own humanity

In a way that she represents as a departure from Kant,[8] Korsgaard says that the strategy of Phase One cannot in itself take us as far as the *moral* law: that is, it cannot establish that the law we must abide by is one that constrains our treatment of others in any recognizably moral way, in either a positive or negative sense, in terms of our obligations to do things for them, or to avoid acting against them. For example, the rational egoist acts on a practical law, in as much as she adopts the principle of always acting to promote her interests, and that seems sufficient to provide the kind of coherent structure to the will that is constitutive of free agency on this

---

Formula of Universal Law. The categorical imperative merely tells us to choose a law. Its only constraint on our choice is that it has the form of a law. And nothing determines what the law must be. *All that it has to be is a law*.'

8 Cf. (Korsgaard 1996b*:* 98-100, 221-222, 233, 237). As she also points out (e.g. p. 99, n. 9), this is also a departure from her earlier self in papers reprinted in (Korsgaard 1996a), where she moves straight from her solution to the Kantian antinomy of self-legislation to the moral law: cf. (pp. 166-7).

account. Korsgaard therefore allows (in a way that a more traditional Kantian might not)[9] that nothing she has said so far establishes that the principle the agent needs to adopt is anything we would recognize as a moral principle, 'the law of what Kant calls the Kingdom of Ends, the republic of all rational beings' (Korsgaard 1996b: 99). To get to *this* law, Korsgaard holds, she must first argue that you must place a value on your own humanity (which is what she does in Phase Two), and then that you must value the humanity of others (which is what she does in Phase Three); once this is established, then the agent cannot adopt the principle of self-interest or any other such non-moral principle as her law, because this would violate the value of humanity. As Korsgaard puts it: '[The] argument…aims to move from the formal version of the categorical imperative to moral requirements by way of the Formula of Humanity' (Korsgaard 1999: 28, n. 23).

On this approach, then, the role of Phase Two is to be a stepping stone to establishing the value of humanity, which is a conclusion of the whole argument in Phase Three; and that stepping stone is to establish to the agent the value of her humanity. To see why Phase Three is required, we can ask how Phase Two falls short: why isn't establishing the value of the agent's own humanity sufficient to lead the agent to adopt the moral law? The answer, of course, is that even if the agent recognizes the value of her own humanity, to be moral she needs to respect the value of others, and it is this shift from agent-relative to agent-neutral reasons that Phase Three is designed to achieve.

Phase Three: From the categorical imperative to the moral law, step 2: the value of humanity in general

9 Cf. (Ginsborg 1998: 8).

As with Phase One, this phase of the argument is not a transcendental one;[10] and it is perhaps the part of Korsgaard's overall approach that has been most brusquely dismissed by critics. Korsgaard's strategy here is to claim that the egoist's agent-relative reasons are private, and then to use considerations from Wittgenstein's private language argument to show that this would make them incoherent as reasons, for this argument shows that reasons must be public to be reasons at all.[11] But Korsgaard's critics have been unimpressed by the suggestion that agent-relative reasons are in fact private ones, in any sense that brings in Wittgensteinian considerations: as Skorupski puts it, 'others can "share" the normative force of the egoist's reasons; that is, they can understand his reasons and, if egoism were right and they were rational, could acknowledge their force (as agent-relative reasons)' (Skorupski 1998: 348-9).[12]

This completes my outline of Korsgaard's overall argument. I will not say anything more about Phase One or Phase Three, and in particular I will not attempt to defend Korsgaard from her critics over the latter, because my focus here is intended to be on the transcendental part of her strategy, which is in Phase Two. It is therefore now time to look in more detail at what the transcendental argument in Phase Two is meant to be.

II

---

10 At least, I don't think it is, and Korsgaard doesn't claim it is either. But for the suggestion it should be seen in this way, see (Skidmore 2002: 135).

11 Cf. (Korsgaard 1996b: 132ff).

12 Skorupski is here echoing Nagel's complaint in his response to Korsgaard in (Nagel 1996: 208); and cf. also (Skidmore 2002: 135-7).

On the account I have given of Phase Two, the aim here is to establish the value of your humanity, as a way of moving to the establishment of the value of humanity in general in Phase Three, in order to show why the law that the agent chooses in Phase One cannot violate the dignity of persons.

To those familiar with debates surrounding transcendental arguments, this may immediately raise concerns. For, it may seem that Korsgaard is straightaway making claims for her transcendental argument that have been famously rendered problematic by Barry Stroud, by using that argument to establish a conclusion about how things are, viz. that your humanity has value. As has been much discussed, Stroud suggested in his 1968 article that such world-directed claims can invariably be resisted by the sceptic, and weakened to appearance or belief directed ones, so that the conclusion of a plausible transcendental argument will only tell us how things must appear to us or how we must believe them to be, in order to make possible thought, language, experience or whatever.[13] If Korsgaard's transcendental argument is making a world-directed claim, therefore, it may seem that this Stroudian worry needs to be addressed.

Of course, if Korsgaard wanted to defend a world-directed transcendental argument, she could perhaps do so by questioning Stroud's reasons for thinking that such arguments can always be weakened by the sceptic to what we must believe or how things must appear. But even if Stroud's position is questionable in general,[14] it may seem that there is something especially problematic in taking Korsgaard's argument in a strong or ambitious form: for it may just seem incredible to think that you could be given an argument to establish that your humanity has value, in a world-directed sense. This incredibility is clearly felt by Skorupski

13 (Stroud 1968: 255-6). For further discussion of Stroud's position, see (Stern 2000).

14 In (Stern 2007b), I argue that Stroud's argument is not as compelling as is usually assumed, but that a better argument can be offered to the same modest effect.

when he writes: 'It would be gratifying to have it demonstrated by pure philosophy that one is important… But in the absence of contagious magic the demonstration seems less than cogent' (Skorupski 1998: 350). The worry here, I think, is hubris: how can it be established that we have value as such, when seen in the scheme of things it seems we have no more significance than anything else—when, as Hume put it, 'the life of man is of no greater importance to the universe than that of an oyster' (Hume 1965: 301). Thus, even if a case could be made for world-directed transcendental arguments in general (contra Stroud et al), is it reasonable to think that such a case can be made concerning my value as a human being?

However, whatever the force of these concerns, it is not clear that they are worries that need apply to Korsgaard. For, she herself does not propose any such ambitious, world-directed transcendental argument,[15] but instead puts forward a modest argument, combined with a kind of anti-realism or perspectivism about value, that does get to the stronger conclusion that your humanity has value, but only where that value is conceived of in this perspectival way.[16]

That Korsgaard sees herself as approaching things this way is clear from what she says when she summarizes her transcendental argument:

---

15 I would therefore contrast Korsgaard's position with Allen Wood's, who follows a similar argument to Korsgaard, but to a more realist conclusion, but without attempting to address Stroudian concerns: cf. (Wood 1999: 125-32); and his review of Korsgaard's *Creating the Kingdom of Ends* in (Wood 1998).

16 Skidmore accepts this defence of Korsgaard's transcendental argument in Phase Two, but as we have seen, thinks Korsgaard also has a transcendental argument in Phase Three, which is more ambitious, and so which falls foul of Stroud's criticisms: see (Skidmore 2002: 134-40). But as I have mentioned, Korsgaard herself doesn't present Phase Three as a transcendental argument, so it is debatable whether these issues apply to it. In her more recent rehearsal of the argument in (Korsgaard 2009: 22-5), she also distinguishes clearly between these phases, and again only talks of the second in transcendental terms.

> The argument I have just given is a transcendental argument. I might bring that out more clearly by putting it this way: rational action exists, so we know it is possible. How is it possible? And then by the course of reflections in which we have just engaged, I show you that rational action is possible only if human beings find their own humanity to be valuable. But rational action is possible, and we are the human beings in question. Therefore we find ourselves to be valuable. Therefore, of course, we are valuable. (Korsgaard 1996b: 123-4)

Korsgaard then goes on:

> You might want to protest against that last step. How do we get from the fact that we find ourselves to be valuable to the conclusion that we are valuable? (Korsgaard 1996b: 124)

And here is Korsgaard's response to this worry:

> [T]here's a good reason why the argument must take this form after all. Value, like freedom, is only directly accessible from within the standpoint of reflective consciousness. And I am now talking about it externally, for I am describing the nature of the consciousness that gives rise to the perception of value. From this external, third-person perspective, all we can say is that when we are in the first-person perspective we find ourselves to be valuable, rather than simply that we are valuable. There is nothing

> surprising in this. Trying to actually see the value of humanity from the third-person perspective is like trying to see the colours someone sees by cracking open his skull. From the outside, all we can say is why he sees them.
>
> Suppose you are now tempted once more to say that this shows that value is unreal just as colour is unreal. We do not need to posit the existence of colours to give scientific explanations of why we see them. Then the answer will be the same as before. The Scientific World View is no substitute for human life. If you think it is unreal, go and look at a painting by Bellini or Olitski, and you will change your mind. If you think reasons and values are unreal, go and make a choice, and you will change your mind. (Korsgaard 1996b: 124-5)

Korsgaard is thus agreeing with the central Humean idea, that from the point of view of the universe nothing really has value; but Korsgaard doesn't claim to be operating from that point of view. Rather, she is engaging with agents who have a perspective on the universe that involves the experience of values and making judgements about them, just as much as they have a perspective that involves experiencing colours and making judgements about them too. So, she holds, if we can establish that from our perspective we must experience or judge that we ourselves have value, that is good enough for this exercise. Korsgaard can therefore agree with the Humean point that to think that we could establish anything more about this world is absurd, and can likewise claim that her transcendental argument doesn't have to be ambitious in this sense, while still insisting that no more than this is required, once this conception of value is accepted.

Now, of course, this approach to the notion of value, and whether it is substantive or realist enough, will be controversial;[17] but I take it here that enough people will find it congenial to serve as an adequate way of allowing Korsgaard to present her transcendental argument in modest terms.[18] We are now in a position to see what that modest transcendental argument is

---

17 In his response to Korsgaard in (Korsgaard 1996b), G. A. Cohen expresses some misgivings on this score: see Cohen (1996: 186); and for a defence of Korsgaard's approach, see (Gibbard 1999: 153): 'One aspect of Korsgaard's argument will be controversial, but not with me. It is transcendental: it takes something we can't act without accepting, derives a consequence, and then embraces the consequence… [But] if the argument goes through as intended, its conclusion doesn't follow logically from its premises—that's the worry. Mightn't it be that although merely acting at all commits us to *thinking* that humanity has value, in fact it doesn't *have* value? Korsgaard, though, say I, has every right to rely on such arguments. Suppose she is right, and in settling whether to act, I've settled whether to believe humanity valuable. I'll then act and voice the conviction to which acting commits me: Humanity is valuable. What other conceivable access can I have, after all, to the question of whether humanity is valuable, but to reflect on what to do? The value of humanity or its lack isn't a feature of nonnatural space, glimpsed by intuition. Thinking humanity valuable, if Korsgaard is right, is an inseparable part of thinking what to do and why. Whether to think humanity valuable is just the question, whether to value humanity'. Cf. also (Bittner 1989: 24): 'Now actually we may disregard the difference between showing that moral demands are valid and showing that they must be considered valid. For the realization that given certain basic features of our lives we cannot help but acknowledge moral demands is tantamount to having their validity demonstrated to us'.

18 Skorupski seems to think that even such a modest transcendental argument is as problematic as a more ambitious one, when the full quotation from him cited earlier runs: 'It would be gratifying to have it demonstrated by pure philosophy that one is important. *Or even—to put it with due Kantian caution—that one must take oneself to be.* But in the absence of contagious magic the demonstration seems less than cogent' [my emphasis]. This is because Skorupski does not see why there might not be valuable things to be done independently of our having value (cf. (Skorupski 1998: 350): 'Even if

meant to be, where all that it intends to establish qua transcendental argument is that we must value our own humanity. I will offer two accounts of this argument, and claim that the second is to be preferred.

III

The first account of the argument I will consider runs as follows:

1. You cannot act unless you can take some impulse to be a reason to act.
2. You cannot take some impulse to be a reason to act unless it conforms to some way in which you identify yourself (a practical identity).
3. You cannot adopt a particular practical identity unless you also adopt humanity as a practical identity.
4. You cannot adopt humanity as a practical identity unless you value your humanity.
5. Therefore, you must value your humanity, if you are to be an agent.

Let me consider in more detail what this all means.

---

humanity is worthless might there not sill be valuable things to be done?'), or how our having value can 'magically' confer value on other things. I think Korsgaard's response, however, would be to say that the alternative is equally 'magical'—namely, how can things have an intrinsic value in themselves? For this sort of worry concerning intrinsic value, cf. also (Street 2008).

As with most transcendental arguments, Korsgaard is trying to begin with a premise that the sceptic can be expected to accept, where here the premise is that he is capable of agency. Now, by this Korsgaard doesn't just mean behaviour or bodily movement, but some exercise of the will. Moreover, she holds, to exercise the will and so act in this way, it is not sufficient that whenever an impulse to act or a desire assails you, you will follow it, for then you are not deciding to act at all. Rather, the way to act is to act for a reason: to decide that this impulse (for example, to buy this toy) is a good one (for example, because it will make my daughter happy).

The second step is to introduce a more distinctively Korsgaardian idea: that if action requires the having of reasons, those reasons are not 'out there' in the world, but come from the way in which doing certain actions would relate to the kind of person you are. Thus, it is qua my daughter's father that I have a reason to buy her this toy, if it would make her happy. As Korsgaard puts it: 'It is necessary to have some conception of your practical identity, for without it you cannot have reasons to act. We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves' (Korsgaard 1996b:120). Unless we had some such practical identity, Korsgaard claims, there would be no reason for us to act on one impulse rather than not, and thus no possibility of rational agency at all.

Obviously one way to resist Korsgaard's argument at this point would be to opt for a more realist conception of reasons, and to claim that for us to have a reason to act on an impulse is just a feature of the situation, independently of whether or not this relates to our practical identity: for example, it is just the potential happiness of my daughter that constitutes a reason for me to obey my impulse to buy the toy. But much of the first two chapters of *The Sources of Normativity*, as well as related papers, is spent arguing against realism of this kind.[19] Korsgaard's position here is complex, and has several strands. One strand is that realism is

19 See (Korsgaard 1996b: 7-89) and (Korsgaard 2003).

unable to explain the felt obligatoriness of moral reasons,[20] where by relating this to the person's sense of self, we can see why she must act, to preserve her sense of who she is.[21] Another strand is that the realist faces a regress of justification or an arbitrary foundationalism, as either one reason is grounded on another, or the regress is brought to a stop by fiat;[22] by contrast, on her position the reasons that apply to the agent can be explained in terms of her practical identity, where (as we shall see) Korsgaard thinks that the regress can be brought to a satisfactory end. Korsgaard thinks that both of these features mean that her position is better placed that the realist's to deal with the moral sceptic, who asks why she should act morally. A final strand in Korgaard's case against realism is an argument from autonomy: if the reasons we

---

20 'According to…realism…there are facts, which exist independently of the person's mind, about what there is reason to do; rationality consists in conforming one's conduct to those reasons… The difficulty with this account in a way exists right on its surface, for the account invites the question why it is necessary to act in accordance with those reasons, and so seems to leave us in need of a reason to be rational… we must still explain why the person feels it *necessary* to act on those normative facts, or what it is about *her* that makes them normative *for her*. We must explain how these reasons get a grip on the agent' (Korsgaard 1997: 240).

21 Cf. (Korsgaard 1996b:100-102).

22 Cf. (Korsgaard 1996b: 33): 'As these arguments show, realism is a metaphysical position in the exact sense criticized by Kant. We can keep asking why: "Why must I do what is right?"—"Because it is commanded by God"—"But why must I do what is commanded by God?"—and so on, in a way that apparently can go on forever. This is what Kant called a search for the unconditioned—in this case, for something which will bring the reiteration of "but why must I do that?" to an end. The unconditional answer must be one that makes it impossible, unnecessary, or incoherent to ask why again. The realist move is to bring this regress to an end by fiat: he declares that some things are *intrinsically* normative… Having discovered that he needs an unconditional answer, the realist straightaway concludes that he has found one.'

have to act are independent of us, then in acting on those reasons we are not acting freely.[23] On the other hand, if reasons stem from our practical identity, then that makes them intrinsic to who we are, and so compatible with our agency.

Now, of course, all these points against realism can be and have been resisted by realists.[24] But perhaps again we can give Korsgaard the benefit of the doubt here, as many would share her conviction that realism is indeed problematic in the ways that she suggests.

Let us move, then, to the third premise of the argument, where having shown that to have a reason to act this must relate to a particular practical identity, Korsgaard tries to show that no particular practical identity can be adopted unless you adopt the practical identity of being human. I think the idea here is as follows.

Suppose you take an impulse like wanting this toy to be a reason to act because it conforms to your particular practical identity of being a father. But as a reflective agent, you can then ask: what reason have I got to adopt this particular practical identity of being a father? You cannot give as a reason: because then I will go around doing good things like buying toys for my daughter, because doing those things are only reasons for you from the perspective of this

---

23 (Korsgaard 1996b: 5): 'If the real and the good are no longer one, value must find its way into the world somehow. Form must be imposed on the world of matter. This is the work of art, the work of obligation, and it brings us back to Kant. And this is what we should expect. For it was Kant who completed the revolution, when he said that reason—which is form—isn't in the world, but is something we impose on it. The ethics of autonomy is the only one consistent with the metaphysics of the modern world, and the ethics of autonomy is the ethics of obligation.'

24 For some notable responses, see (Gaut 1997), (Regan 2002), (Fitzpatrick 2006), (Crisp 2006: 49-56), (Parfit 2006), (Wallace 2006: 71-81). I consider Korsgaard's argument from autonomy further in (Stern 2007a).

identity, which is precisely what is in question. And you cannot give as a reason some further particular practical identity, like being a husband, because the same questions can be raised about that. If you are to halt the regress of reasons, therefore, you must appeal to some reasons that are not grounded in any particular practical identity and so an identity that is likewise not so grounded or 'conditioned': and the only identity of which this is true is the identity of humanity, for without that identity, you could not see yourself as an agent with reasons at all, so it itself does not rest on any reasons given to it by some further particular practical identity. As Korsgaard puts this, beginning with the passage we quoted earlier covering premises 1 and 2:

> It is necessary to have some conception of your practical identity, for without it you cannot have reasons to act. We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves. Yet most of the self-conceptions which govern us are contingent… What is not contingent is that you must be governed by some conception of your practical identity. For unless you are governed by some conception of your personal identity, you will lose your grip on yourself as having any reason to do one thing rather than another—and with it, your grip on yourself as having any reason to live and act at all. But this reason for conforming to your particular practical identities is not a reason that springs from one of those particular practical identities. It is a reason that springs from your humanity itself, from your identity simply as a human being, a reflective animal who needs reasons to act and to live. (Korsgaard 1996b: 121)

I think we can view Korsgaard's position here this way. My practical identity is my sense of who I am. As a reflective agent, I can see that I could be brought to give up any particular practical identity I may have, such as being a father, husband, Englishman, university lecturer etc., as I come to see that they are not really essential to me as such, in a way that 'alienates' me from

them. But I cannot give up my sense of being a person who can think about who he is in the same way, because to do this I would have to be thinking about who I am—and it is this which Korsgaard thinks is distinctive of the practical identity of humanity. In this case, therefore, no more basic identity can supply me with a reason to adopt this identity, any more than some more basic logical principle can supply me with a reason to believe the principle of non-contradiction, in which case I am entitled to treat it as just the sort of 'unconditioned' stopping point that is able to bring the regress of reasons to a principled end.

Now let's consider the final premise: you cannot adopt humanity as a practical identity unless you value your humanity. Thus far, Korsgaard has shown that you are required to adopt humanity as your practical identity if you are to adopt any practical identity at all, as this identity brings a halt to the regress. But suppose you didn't value your identity as human—meaning here being a reflective agent—but just saw it merely as a necessary fact about yourself, which nonetheless you felt neutral about, or even rather regretted and despised?[25] In this case, however, humanity would not bring a halt to the regress, because unless you saw humanity as valuable, it could not give any reasons to act in itself or to adopt any other particular practical identity, where unless it does so, you cannot continue to think of yourself as a rational agent with reasons. It is therefore a necessary condition of having such reasons that you value your humanity: according to Korsgaard, the price of denying this is to see yourself as living in a world in which there are no reasons to act and thus no way to be an agent at all. In a sense, Korsgaard admits, she cannot prevent the sceptic paying the price if he is determined to do so, in a kind of suicidal abandonment of agency;[26] but in another sense we have little choice but to see

25 Cf. (Gibbard 1999: 154): 'Why, though, couldn't I think of reflective choice as a burden, only mitigated by some admirable way that people like me handle it?… couldn't I still disvalue the sheer state of being a reflective chooser?'

26 Cf. (Korsgaard 1996b: 160-4).

ourselves as agents, and so (she thinks) to accept the conclusion of this part of her argument, that you must value your humanity.

IV

Let us call this version of Korsgaard's argument the regress of identities argument. I now want to consider an objection to it.[27]

This concerns whether Korsgaard is right to think that if we work merely with our particular practical identities, we will be threatened with a regress. Now Korsgaard admits, of course, that as a matter of psychology some of us may not reflect on our particular practical identities, and so not face the regress in our daily lives; but that, she argues, just shows that we can be insufficiently reflective and doesn't show we ought not to feel the regress.[28] But, why ought we to feel the regress? Korsgaard's idea seems to be that our particular practical identities are contingent, and that we could therefore always be brought to give them up by finding other identities that are more compelling:

> You may cease to think of yourself as a mother or a citizen or a Quaker… This can happen in a variety of ways: it is the stuff of drama, and perfectly familiar to us all. Conflicts that arise between identities, if sufficiently pervasive or severe, may force you to give one of them up: loyalty to your country and its cause may turn you against a

27 I also have my doubts about the step from (4) to (5), but will leave these aside for now.

28 Cf. Korsgaard's comments on the member of the Mafia brought into the discussion by G. A. Cohen: see (Korsgaard 1996b: 257-8).

> pacifist religion, or the reverse. Circumstances may cause you to call the practical importance of an identity into question: falling in love with a Montague may make you think that being a Capulet does not matter after all. Rational reflection may bring you to discard a way of thinking of your practical identity as silly or jejune. (Korsgaard 1996b: 120)

Now Korsgaard is of course correct that this sort of thing can occur, and when it does, in a way that should lead you to question your particular practical identity. But Korsgaard seems to think that she can generalise from this, to show that it can apply to every particular practical identity for every agent. But I think that the kind of self-doubt that Korsgaard sees as pervasive is in fact harder to achieve than she realises—where I don't just mean psychologically harder (which I agree with her is irrelevant here), but rationally or normatively. I think one can see this by looking at her examples. It is important to the persuasiveness of her examples, that the competing identity is not just different from the one you hold, but can itself supply reasons in its favour: for example, the pacifist can nonetheless come to see some value in fighting for his country. But might there not be particular practical identities which nonetheless seem invulnerable to competing reasons in this way, despite being identities we accept are contingent, in the sense that we can see we might not have had the identity in question? For example, suppose that I am a loving son. I am certainly conscious that I might not have been, had I been raised in a different way or in a different time or place, so it is a contingent identity in that sense. But suppose someone offers me an alternative identity. Unless they can supply me with reasons to think being a loving son is wrong or mistaken in some way, why should I be brought to doubt my identity? But it might seem that the only reasons that would count as reasons to give up that identity are ones that would only do so once that identity has been given up. So, for example, you might say I should become a ruthless city trader, and so stop wasting my time visiting my sick mother and spend it arranging profitable deals instead. But what

reason could you give me for taking this seriously as an identity, given that you are asking me to betray everything I hold dear? It seems that you would need to give me some internal grounds for giving up my identity (for example, that my loving regard for my mother is making her life worse and not better), which even in the case of an identity which is contingent may not be forthcoming. Or you would need to appeal to common ground between being a son and the values that leads to, and being a ruthless city trader, just as the values of the pacifist and non-pacifist may be said to coincide at certain points, from which the divergences can be explored. But in so far as this is possible, then we can continue to operate with reasons at the level of converging particular practical identities, rather than moving to the kind of universal and necessary identity that Korsgaard thinks is required. So, if this is right, Korsgaard has arguably not done enough to show that it is only the identity of humanity that is invulnerable to the regress issue and hence 'unconditioned', and not also certain particular practical identities; and if this is right, her transcendental argument in this form collapses at step (3).[29]

V

I now want to look at a second transcendental argument that can be found in Korsgaard's work, which I think fares better than the first one.

This second argument is modelled on an argument that Korsgaard finds in Kant, and which she outlines as follows:

29 For similar sorts of objections, see (Schneewind 1998: 43-8). See also Korsgaard's reply in (Korsgaard 1998).

> [Kant] started from the fact that when we make a choice we must regard its object as good. His point is the one I have been making—that being human we must endorse our impulses before we can act on them. He asked what it is that makes these objects good, and, rejecting one form of realism, he decided that the goodness was not in the objects themselves. Were it not for our desires and inclinations—and for the various physiological, psychological, and social conditions which gave rise to those desires and inclinations—we would not find their objects good. Kant saw that we take things to be important because they are important to us—and he concluded that we must therefore take ourselves to be important. In this way, the value of humanity itself is implicit in every human choice. If complete normative scepticism is to be avoided—if there is such a thing as a reason for action—then humanity, as the source of all reasons and values, must be valued for its own sake. (Korsgaard 1996b: 122)[30]

This argument can be laid out as follows:

1. To rationally choose to ϕ, you must regard ϕ-ing as good.

2. You cannot regard ϕ-ing as good in itself, but can only regard ϕ-ing as good because it satisfies your needs, desires, inclinations, and so on.

30 Korsgaard claims to take her inspiration from Kant's *Groundwork of the Metaphysics of Morals* (4:427-8), which she discusses further in her paper 'Kant's Formula of Humanity', reprinted in (Korsgaard 1996a: 106-32). I leave aside here whether or not Korsgaard is correct in her reading of Kant; but for some doubts on this score, see (Timmermann 2006).

3. You cannot regard your desiring or needing to ϕ as making it good unless you regard yourself as valuable.

4. Therefore, you must regard yourself as valuable.

Consider this example. To rationally choose to eat this piece of chocolate cake, I must think that the cake is good in some way. How can I regard it as good? It seems implausible to say that the cake is good in itself, of intrinsic value. It also seems implausible to say that it is good just because it satisfies a desire as such: for even if I was bulimic it might do that, but still not be regarded as good. A third suggestion, then, is that it can be seen as good because it is good for me, as satisfying a genuine need or desire of mine. But if I think this is what makes the piece of cake good, I must value myself as, otherwise, I could not hold that satisfying me is sufficient to make something good enough for it to be rational for me to desire it; so I must regard myself as valuable. Put conversely: suppose that you thought that you and your life were utterly worthless, pointless, meaningless—that in your eyes, you were valueless. And suppose that you are faced with a piece of cake: on what basis would you chose it eat it? It seems unlikely that there is something intrinsically good about eating it, or that you should do so just because you find yourself with a desire to do so, even while finding your existence valueless. It seems that the only reason to do so would be if you thought eating the cake brought you some genuine benefit—but if you thought your life was worthless, how could you see this as a reason either? Why is bringing benefit to something that in your eyes is so utterly without value a reasonable thing to do?

There are some dangers in this argument, however. One, which Korsgaard considers, is that it might lead to 'self-conceit':[31] that is, I might conclude from this that I am supremely

31 (Korsgaard 1998: 54). Cf. also (Korsgaard 1996b: 249-50).

valuable, simply as Bob Stern, which could obviously then get in the way of my ethical treatment of others. But, this worry might be lessened by the thought that while the argument gets me to see that I must find something valuable about me, it need not be anything about me in particular, and perhaps could instead be something about me that is more general—such as my humanity or personhood. However, while Korsgaard says that reflection will indeed lead us in this more general direction, we will need to see how. A second, perhaps related, worry is that this argument has a troubling parallel in the case of Satan, where Satan goes through 1 to 4 above, and concludes that he must regard his devilish nature as valuable. If this argument somehow entitles us to regard our humanity or personhood as valuable, why doesn't it entitle Satan to think the same about his nature? This is not the same as self-conceit, because he is not valuing himself as Satan just qua Satan; he is valuing his nature, just as we are valuing ours. Nor does devilishness seem any less central to his nature than humanity is to ours. So it is hard to see how the Satanic parallel can be avoided by the argument as it stands.

Nonetheless, it is possible that something can be built on the central idea of the argument, which I take to be this: As long as we think we can act for reasons based on the value of things, but at the same time reject any realism about that value applying to things independently of us, then we must be treated as the source of value and in a way that makes rational choice possible. We can therefore see Korsgaard's second argument as attempting something along these lines, using her notion of practical identity to perhaps avoid the two problems we have identified with the Kantian argument.

Here, then, is an outline of Korsgaard's second argument:

1. To rationally choose to ϕ, you must take it that ϕ-ing is the rational thing to do.

2. Since ϕ-ing in itself gives you no reason to ϕ, you can take it that ϕ-ing is the rational thing to do only if you regard your practical identity as making it rational to ϕ.

3. You cannot regard your practical identity as making ϕ-ing the rational thing to do unless you can see some value in that practical identity.

4. You cannot see any value in any particular practical identity as such, but can regard it as valuable only because of the contribution it makes to giving you reasons and values by which to live.

5. You cannot see having a practical identity as valuable in this way unless you think your having a life containing reasons and values is important.

6. You cannot regard it as important that your life contain reasons and values unless you regard your leading a rationally structured life as valuable .

7. You cannot regard your leading a rationally structured life as valuable unless you value yourself qua rational agent.

8. Therefore, you must value yourself qua rational agent, if you are to make any rational choice.

The first step is now familiar: To act is to do or choose something for a reason. The second step is also now familiar: Korsgaard thinks that we have reasons to act because of our practical identities, not because acts have reasons attached to them in themselves. Once again, realists might demur,[32] claiming that some actions are rational things to do, because some things have value as such: so, perhaps knowledge is valuable in itself, thereby making it rational to seek it.[33] But as before, let us leave such worries aside and assume with Korsgaard that nothing is objectively rational for us to do.

---

32 See, for example, (Scanlon 1998: 55-72), and (Kerstein 2002: 70-2).

33 Cf. (Regan 2002: 272).

The third step asks how a practical identity can make something into a reason for an agent: how can the fact that I am a father make it rational for me to buy my daughter this toy? The thought here is that it can only do so if I see value in that identity. Korsgaard stresses this when she writes:

> The conception of one's identity in question here is not a theoretical one, a view about what as a matter of inescapable scientific fact you are. It is better understood as a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking. (Korsgaard 1996b: 101)

So, being a father, whether contingently or essentially, gives one no reason to be a caring or devoted father of a sort that would have good reason to buy a daughter a gift; rather, valuing one's fatherhood does this.

But (moving on to step 4), how can I see my particular practical identity as valuable? I think Korsgaard's position here is that I cannot see any value in any particular practical identity as such, and this might seem to rest on something like the sort of regress argument we looked at and criticised above: one will always be faced with requiring a reason for valuing any particular practical identity, and this regress cannot be halted by any such identity. But, as we saw, this thought can perhaps be resisted. However, I think that even without a regress argument, Korsgaard can make her point here, more by using the objections to realism considered previously: namely, that to see value in any particular practical identity as such is to be committed to realism, to thinking that being a father, an Englishman, a university lecturer or whatever matters as such; or (in a way that is in the end equally realist), it matters because of the intrinsically valuable things it leads you to do. But, as we have seen, Korsgaard also takes

such realist positions to be problematic, so can perhaps use such arguments here, without appealing to the regress considerations at all.

So suppose we allow that no particular practical identity can be seen to have value in itself; Korsgaard then offers as the only remaining explanation of its value to the agent that has that identity, that such identities have the general capacity of enabling the agent to live a life containing reasons: because I have whatever particular practical identities I do (father, Englishman, university lecturer…), I can then find things to be valuable and act rationally accordingly, in a way that gives me unity as a subject. As Korsgaard puts it: 'To be a thing, one thing, a unity, an entity; to be anything at all: in the metaphysical sense, that is what it means to have integrity. But we use the term for someone who lives up to his own standards. And that is because we think that living up to them is what makes him one, so what makes him a person at all' (Korsgaard 1996b:102).

But then (step 5), to think that this makes having some sort of particular practical identity important, you must think that it matters that your life have the sort of rational structure that having such identities provides; but (step 6), to see that as mattering, you must see value in your leading a rationally structured life. And then, finally, to see value in your leading such a life, you must see your rational nature as valuable, which it to value your humanity.

Does this Korsgaardian argument avoid the pitfalls of the Kantian one discussed earlier? I think it avoids the problem of self-conceit, because it does seem that what you end up valuing is not yourself simply as such, but yourself qua rational agent. And I think as I have presented it, it avoids the problem of the Satanic parallel, because all it shows is that Satan must value his rational nature, not his devilishness.

For both these problems to be avoided, however, it is important to run the argument as I have done, not as it is sometimes presented by Korsgaard, which is via the notion of need.[34] This would follow the same premises as before for 1-5, and then go as follows:

6.* You cannot regard it as important that your life contain reasons and values unless you take your need to lead this sort of life as important.

7.* You cannot take this need to be important unless you take yourself to be valuable.

8.* Therefore, you must value yourself, if you are to make any rational choice.

The difficulty with 6*-8*, I think, is that 8* does not stipulate what it is about yourself that you are required to value, so that this could be my sheer particularity (self-conceit), or if I am not in fact human, my non-human nature (Satan). This is because 6* just identifies a need, and says that this need could not be important unless the agent who has the need were seen to be valuable somehow—whereas the previous argument narrows value down to rational agency, and so rules out both self-conceit and devilishness.

---

34 Cf. (Korsgaard 1996b:121): 'Most of the time, our reasons for action spring from our more contingent and local identities. But part of the normative force of those reasons spring from the value we place on ourselves as human beings who need such identities. In this way all values depends on the value of humanity; other forms of practical identity matter in part because humanity requires them'; and (Korsgaard 1996b:125): 'Our other practical identities depend for their normativity on the normativity of our human identity—on our own endorsement of our human need to be governed by such identities—and cannot withstand reflective scrutiny without it. We must value ourselves as human.'

VI

I have therefore reconstructed that part of Korsgaard's strategy which offers an argument to the effect that you must value your humanity, as a transcendental argument. It turns out that if it is to be made plausible in this way, a lot depends on accepting Korsgaard's arguments against realism; but then, many have suspected that some commitment to anti-realism is required to make a transcendental argument convincing. A worry then is that it can appear to make the argument redundant in the standard anti-sceptical case, because anti-realism appears sufficient as a response to scepticism on its own;[35] but in this ethical case, this does not seem to be the issue, so that here this worry is less of a concern. Of course, as Korsgaard herself allows, this transcendental argument in itself is not meant to be sufficient to complete her project, which still requires a third phase, which I have not considered, and which may still be found to be problematic.[36] Nonetheless, I hope to have shown something, which is to have established how

35 For more discussion of this issue, see (Stern 2000: 49-58).

36 There is another objection which I also cannot address in this paper, and which is not considered directly by Korsgaard: namely that if 'valuing your humanity' comes down to 'valuing your rational agency', will that cover enough of what morality is usually meant to cover? Allan Gibbard puts the point this way: 'If valuing my humanity is taking pride in being a reflective chooser, how does that constrain what I do, what I reflectively choose? Perhaps I must nurture my powers of reflective agency… Still, if that's all we must do as human beings, then enlightenment morality is far too narrow: we'll need to oppose pain and seek fulfilment and enjoyment only when they affect our powers of reflective agency' (Gibbard 1999: 156). I am not sure the Kantian would in fact see much of a worry here. Perhaps a deeper concern, however, is that even if the argument I have considered works, there is a problem with Korsgaard's whole *strategy* of attempting to respond to moral scepticism in the way she does, as in the end the reason to be moral becomes grounded in the interest we have in being

the transcendental argument of the second phase can be seen to work, and how it is more plausible than many of Korsgaard's critics have found it; and this I think is an achievement of sorts.[37]

References

Bittner, R. 1989. *What Reason Demands*. Cambridge: Cambridge University Press.

Cohen, G. A. 1996. Reason, Humanity, and the Moral Law. In Korsgaard 1996b: 167-88.

Crisp, R. 2006. *Reasons and the Good*. Oxford: Oxford University Press.

Fitzpatrick, W. J. 2006. The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity. *Ethics* 115: 651-91.

Gaut, B. 1997. The Structure of Practical Reason. In G. Cullity and B. Gaut, eds. *Ethics and Practical Reason*: 161-88. Oxford: Oxford University Press.

---

agents, which is to distort the nature of genuine moral action which should not be grounded in anything *outside* morality itself. I discuss this issue further in (Stern 2010).

37 I gave a first version of this paper at one of the *Transcendental Philosophy and Naturalism* workshops, under the AHRC funded project run by Mark Sacks. Mark offered very useful and characteristically generous comments on that occasion, the last on which I was to see him. I am therefore particularly grateful to Joel Smith and Peter Sullivan for undertaking the publication of this paper, as what is very sadly the last chapter in the discussion that Mark and I had on matters transcendental over many years, from which I learned so much.

Gibbard, A. 1999. Morality as Consistency in Living: Korsgaard's Kantian Lectures. *Ethics* 110: 140-64.

Ginsborg, H. 1998. Korsgaard on Choosing Nonmoral Ends. *Ethics* 109: 5-21.

Hume, D. 1965. Of Suicide. In A. MacIntyre, ed. *Hume's Ethical Writings*: 297-306. New York: Collier Books.

Illies, C. 2003. *The Grounds of Ethical Judgement*. Oxford: Oxford University Press.

Kant, I. 1998. *The Groundwork of the Metaphysics of Morals*. Translated and edited by M. Gregor. Cambridge: Cambridge University Press.

Kerstein, S. J. 2002. *Kant's Search for the Supreme Principle of Morality*. Cambridge: Cambridge University Press.

Korsgaard, C. M. 1996a. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

— 1996b. *The Sources of Normativity.* Cambridge: Cambridge University Press.

— 1997. The Normativity of Instrumental Reason. In G. Cullity and B. Gaut, eds. *Ethics and Practical Reason*: 215-54. Oxford: Oxford University Press. Reprinted in Korsgaard 2008: 27-68.

— 1998. Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind. *Ethics* 109: 49-66.

— 1999. Self-Constitution in the Ethics of Plato and Kant. *The Journal of Ethics* 3: 1-29.

— 2003. Realism and Constructivism in Twentieth-Century Moral Philosophy. *APA Centennial Supplement to the Journal of Philosophical Research*: 99-122. Reprinted in Korsgaard 2008: 302-26.

— 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (Oxford: Oxford University Press).

— 2009. *Self-Constitution*. Oxford: Oxford University Press.

Nagel, T. 1996. Universality and the Reflective Self. In Korsgaard 1996b: 200-09.

Parfit, D. 2006. Normativity. In R. Shafer-Landau, ed. *Oxford Studies in Metaethics*, vol. 1: 325-380. Oxford: Oxford University Press.

Regan, D. H. 2002. The Value of Rational Nature. *Ethics* 112: 267-91.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schneewind, J. B. 1998. Korsgaard and the Unconditional in Morality. *Ethics* 109: 36-48.

Skidmore, J. 2002. Skepticism about Practical Reason: Transcendental Arguments and their Limits. *Philosophical Studies* 109: 121-41.

Skorupski, J. 1998. Rescuing Moral Obligation. *European Journal of Philosophy* 6: 335-55.

Stern, R. 2000. *Transcendental Arguments and Scepticism*. Oxford: Oxford University Press.

— 2007a. Freedom, Self-Legislation and Morality in Kant and Hegel: Constructivist vs Realist Accounts. In E. Hammer, ed. *German Idealism: Contemporary Perspectives*: 245-66. London: Routledge.

— 2007b. Transcendental Arguments: A Plea for Modesty', *Grazer Philosophische Studien* 74: 143-61.

— 2010. Moral Scepticism and Agency: Kant and Korsgaard? *Ratio* 23: 453-74.

Street, S. 2008. Constructivism about Reasons. In R. Schafer-Landau, ed. *Oxford Studies in Metaethics*, vol. 3: 207-46. Oxford: Oxford University Press.

Stroud, B. 1968. Transcendental Arguments. *Journal of Philosophy* 65: 241-56.

Timmermann, J. 2006. Value Without Regress: Kant's 'Formula of Humanity ' Revisited. *European Journal of Philosophy* 14: 69-93.

Wallace, R. J. 2006. Normativity and the Will. Reprinted in his *Normativity and the Will*: 71-81. Oxford: Oxford University Press.

Wood, A. W. 1998. Review of Korsgaard *Creating the Kingdom of Ends*. *Philosophical Review* 107: 607-611.

— 1999. *Kant's Ethical Thought*. Cambridge: Cambridge University Press.