Ethical Theory and Moral Practice (2018) 21:887–902
https://doi.org/10.1007/s10677-018-9925-3

Korsgaard's Other Argument for Interpersonal Morality: the Argument from the Sufficiency of Agency

Sem de Maagt[1]

Accepted: 21 August 2018 / Published online: 25 August 2018

Abstract
Christine Korsgaard's (1996, 2009) argument for the claim that one should not only value one's own humanity but also the humanity of all other persons, 'the publicity of reasons argument', has been heavily criticized and I believe rightly so. However, both in an early paper (1986) and in her most recent work (forthcoming), Korsgaard does not rely on controversial, Wittgensteinian ideas regarding the publicity of reasons, but instead she uses a different argument to justify interpersonal morality, which I will refer to as 'the argument from the sufficiency of agency'. The goal of this paper is to evaluate whether the argument from the sufficiency of agency can succeed where the publicity of reasons argument fails. I will argue that although the argument from the sufficiency of agency is potentially more promising, it fails to justify a categorical and universal principle of interpersonal morality. I argue, however, that this failure has less to do with the argument from the sufficiency of agency itself and more with Korsgaard's specific version of it. This leaves open the possibility that other Kantian constructivist arguments from the sufficiency of agency might be more successful.

Keywords Kantian constructivism · Transcendental arguments · Korsgaard · Value · Humanity

1 Introduction

In recent years, Christine Korsgaard has embarked on the ambitious ethical project to provide a Kantian constructivist argument for a categorical and universal principle of interpersonal morality, according to which we must value the humanity of all persons. According to Korsgaard, her argument is ultimately meant to show "that Enlightenment morality is true" (Korsgaard 1996, 123). Korsgaard's argument consists of two steps. In the first step, she employs a transcendental argument to show that any agent necessarily

✉ Sem de Maagt
s.demaagt@uu.nl; semdemaagt@gmail.com

[1] Department of Philosophy and Religious Studies, Utrecht University, Janskerkhof 13A, 3512 BL Utrecht, The Netherlands

has to value her own humanity (1996, 123). Korsgaard, however, stresses that if this argument is successful it "shows only (or at most) that you must place a value on your own humanity, but not yet that you therefore have obligations to other human beings" (Korsgaard 1996, p. 130). The second step of the argument is therefore supposed to show that we do not only have to value our own humanity but also the humanity of all other human beings. In other words, the second step of the argument is supposed to justify conclusions of universal, interpersonal morality.

Both in *The Sources of Normativity* (1996) and more recently in *Self-Constitution* (2009), Korsgaard tries to argue for the latter, most controversial step of her argument by trying to show that reason are what she calls "public in their very essence" (Korsgaard 1996, 135), meaning that "to act on a reason is already, essentially, to act on a consideration whose normative force may be shared with others" (Korsgaard 1996, 136). This Wittgensteinian inspired 'publicity of reasons argument' has been heavily criticized (Skorupski 1998; LeBar 2001; Gert 2002; Skidmore 2002; Wallace 2009; Beyleveld 2015), and I believe rightly so. Elsewhere, I have argued that her slightly revised argument for the publicity of reasons in *Self-Constitution* (2009) fails as well (De Maagt 2018).

However, both in an early paper (1986) and in her most recent work (forthcoming), Korsgaard does not rely on controversial, Wittgensteinian ideas regarding the publicity of reasons, but instead she uses a different, potentially less controversial argument for interpersonal morality, which I will refer to as the 'the argument from the sufficiency of agency'.[1] The basic idea of this argument is that because one's agency is a sufficient condition of one's own humanity having value, one is necessarily committed to the conclusion that *any* agent's humanity, and not just your own humanity, has value. If this argument were successful, there would be no need for Korsgaard to refer to the controversial idea of the publicity of reasons in the first place.

The goal of this paper is therefore to evaluate whether the argument from the sufficiency of agency can succeed where the publicity of reasons argument fails. In contrast to the common practice of Korsgaard's critics to discuss the two steps of Korsgaard's argument independently from each other, I will argue that the success of the argument from the sufficiency of agency crucially depends on the specific interpretation and success of the first step of Korsgaard's argument (the argument for the value of one's own humanity).

I shall argue that Korsgaard's argument taken as a whole fails to lead to moral obligations to others. This failure has, however, less to do with the argument from the sufficiency of agency itself and more with the details of the first step of her argument. The problem with this argument is that it does not lead to any normative conclusions: there is little if nothing that one can do to fail to respect either one's own humanity or, when this argument combined with the argument from the sufficiency of agency, to fail to respect the humanity of others. Although this paper is thus ultimately sceptical about Korsgaard's argument for universal moral obligations towards others, it leaves open the possibly that other Kantian constructivists arguments which rely on the argument from the sufficiency of agency, such as Gewirth (1978), are more successful.

The structure of the paper is as follows: first, I introduce Korsgaard's argument from the sufficiency of agency and I briefly discuss how this argument relates to her publicity of reasons argument (Section 2). I then critically evaluate Korsgaard's attempt to justify interpersonal morality through the argument from the sufficiency of agency by combining

[1] I take this label from a similar argument put forward by Alan Gewirth (1978, 110).

this argument with the argument for the value of your own humanity. I will do so by focussing on Robert Stern's (2011) recent reconstruction and defence of Korsgaard's transcendental arguments for the value of your own humanity, which I take to be strongest version of this argument (Section 3). Finally, I argue that even if we assume that both the argument from the sufficiency of agency and the argument for the value of your own humanity are valid, the two arguments combined neither justifies moral obligations towards oneself (Section 4) nor towards others (Section 5).

2 The Argument from the Sufficiency of Agency

There are roughly two ways in which Kantian constructivists have tried to argue for a universal and categorical principle of interpersonal morality. Arguments from the first-person perspective hold that it is possible to argue from the claim that A must accept whatever are the necessary preconditions of A's agency (e.g. to value A's humanity) to the claim that A is required to assign the same moral status to the necessary preconditions of the agency of other individuals (e.g. to value the humanity of other agents as well). Some think, however, that it is impossible to argue from the necessary preconditions of individual agency to interpersonal morality. This has given rise to various arguments from the second-person perspective which propose to take as their starting point A's involvement in some form of interaction, such as communication, argumentation or shared action, and subsequently explore the necessary conditions of the possibility of these forms of interaction.

Korsgaard's (1996, 2009) widely discussed and criticized publicity of reasons argument is an influential example of an argument from the second person (other examples include Apel 1980; O'Neill 1986; Habermas 1990; Darwall 2006).[2] In *The Sources of Normativity,* Korsgaard explicitly criticises neo-Kantian attempts to justify interpersonal morality by an argument from the first person. Korsgaard puts forward her objection to these arguments by making a distinction between what she calls 'private reasons' and 'public reasons' (Korsgaard 1996, 133). According to Korsgaard, an agent has a *private* reason if a reason has normative force for him or her. An agent has a *public* reason if the reason also has normative force for others, e.g. when your reason to value your own humanity is also a reason for others to value your humanity. Korsgaard argues that the problem with arguments from the first person is that they try to argue from private reasons to public reasons: an argument from the first person tries to show that I need to take *your* reasons into account on the basis of the fact that I need to take *my own* reasons into account. This first-personal strategy, however, does not work:

> Consistency can force me to grant that your humanity is normative for you just as mine is normative for me. It can force me to acknowledge that your desires have the status of

[2] In fact, Korsgaard's (1996, 2009) argument might be best characterized as a 'hybrid' transcendental argument. As I have mentioned above, her arguments consists of two steps. In the first step of her argument, she employs a first-personal transcendental argument in order to argue for the conclusion that agents necessarily have to value their own humanity. In the second step of her argument, she introduces the second-personal idea that all reasons are essentially social reasons and argues that only the social nature of reasons can generate interpersonal morality. In a review paper on Darwall's second-person standpoint, Korsgaard emphasizes the hybrid nature of her position when she claims that "Darwall characterizes me both as someone who thinks all reasons are second-personal and also as someone who thinks that 'moral obligations can be grounded in the constraints of first-personal deliberation alone'. That may sound paradoxical but it is basically right" (Korsgaard 2007, 10).

> reasons for you, in exactly the same way that mine do for me. But it does not force me to share in your reasons, or *make your* humanity normative for me. It could still be true that I have my reasons and you have yours, and indeed that they leave us eternally at odds (Korsgaard 1996, 134).[3]

Korsgaard thus claims that although an argument from the first person could show that *every* agent necessarily has to value his or her own humanity, it does not (and cannot) follow that agents have to value each other's humanity or whatever are the necessary preconditions of agency. Korsgaard thus suggests that a value might be *universal* (in the sense that every agent has to value her own humanity), but not (yet) *public* (in the sense that agents have to value each other's humanity as well). In other words, Korsgaard claims that an argument from the first person cannot show that (moral) egoism is inconsistent. The Wittgensteinian inspired publicity of reasons argument is supposed to show that reasons are essentially public so that there is no need to bridge the gap between private reasons, and public reasons, or between the personal and the interpersonal. But, as already noted above, there is widespread agreement that this publicity of reasons argument does not work, or at least insofar as the argument is supposed to lead to the conclusion that we have to value the humanity of all other persons (Skorupski 1998; LeBar 2001; Gert 2002; Skidmore 2002; Wallace 2009; Beyleveld 2015).

However, both in an early paper (1986) and in her most recent work (forthcoming), Korsgaard does not refer to the publicity of reasons argument, but she instead seems to rely on a first-personal strategy to justify interpersonal morality.[4] In an early paper, 'Kant's Formula of Humanity' (1986), Korsgaard argues that the claim that you must value your humanity commits you to valuing the humanity of *all other* rational agents at the same time. She writes that "if you view yourself as having a value-conferring status [i.e. if your humanity is valuable] in virtue of your power of rational choice, you must view anyone who has the power of rational choice as having, in virtue of that power, a value-conferring status" (Korsgaard 1986, 196). Korsgaard thus claims that anyone who has the power of rational choice, i.e. every agent, must value his or her humanity and that this person must also value the humanity of all other agents.

Unfortunately, however, in this paper Korsgaard says very little about why one must value the humanity of anyone with the power of rational choice. But in a recent paper, Korsgaard (forthcoming) elaborates on her earlier suggestion. She compares the reason for being committed to valuing the humanity of others with extending voting rights to all citizens of a country: if you have voting rights simply by virtue of being a citizen, you should accept that every citizen has voting rights:

> [S]uppose you ask me, "In virtue of what do you have the right to vote here?" and I reply, "I am a citizen of this nation." Citizenship, as I understand it, is a form of normative standing: it gives its possessor certain normative or moral powers. You might reply, "Well, I am a citizen too, so I have the right to vote here as well." Notice that it would not make sense for me to respond, "no, *my own* citizenship has that normative implication – but so far as I am concerned, yours does not" (forthcoming, 28–29).

[3] Korsgaard basis this argument on Williams (1986, 61). For a defence of Gewirth's first-personal argument against Williams' criticism, see Beyleveld (2013).

[4] At the end of this section, I briefly compare the argument from the sufficiency of agency with her characterization of the first-personal strategy in *The Sources of Normativity*.

Korsgaard claims that a similar argument could be made about the value of humanity. The only difference is that whereas the person in the quote above has voting rights by virtue of being a citizen, one's humanity has value by virtue of one being an agent:

> Kant's argument for the Formula of Humanity, treats humanity, or the power of rational choice, as if conferred a kind of normative standing on us. When we look at the argument this way, Kant asks, "in virtue of what do we have the right to treat our ends as good, that is, to confer normative value on them, and so in effect to legislate values?" and he answers, "Our humanity." So the argument assigns us a normative standing in virtue of our humanity, like the normative standing we have in virtue of say, being born in a certain country (Korsgaard forthcoming, 29).

Korsgaard concludes that the fact that we have a normative standing by virtue of the power of rational choice "commits us to assigning the same standing to every other rational being" (Korsgaard forthcoming, 31).

I take Korsgaard to be making the following first-personal argument: if you necessarily have to value your own humanity simply by virtue of being an agent (i.e. having the power of rational choice), you must also value the humanity of any other agent. Likewise, other agents have to value not just their own humanity but also the humanity of others. The reason for this is that being an agent is a sufficient condition of having value, just like citizenship is a sufficient condition of having voting rights. Thus, if being an agent is a sufficient condition of having value, one is committed to the conclusion that any agent, and not just the particular agent you are, has value.

Note that the conclusion of Korsgaard's argument from the sufficiency of agency is thus not merely that one is committed to the claim that every other agent has to value his or her own humanity, but that you, simply by virtue of being an agent, also have to value the humanity of other agents. In this argument, there seems to be no gap between universality and publicity, which the Wittgenstein publicity of reasons is supposed to cross.

The argument from the sufficiency of agency seems much less controversial than the argument from the publicity of reasons. Alan Gewirth (1978), who puts forward a very similar argument, argues that the argument from the sufficiency of agency is simply an application of 'the criterion of relevant similarities', which goes as follows:

> [I]f some predicate P belongs to some subject S because S has the property Q (where the 'because' is that of sufficient reason of condition), then P must also belong to all other subjects S1, S2,..,Sn that have Q. If one denies this implication in the case of some subject, such as S1, that has Q, then one contradicts oneself. For in saying that P belongs to S because S has Q, one is saying that having Q is a sufficient condition of having P; but in denying this in the case of S1, one is saying that having Q is not a sufficient condition of having P (Gewirth 1978, 105).

The criterion of relevant similarities thus states that if Q is a sufficient condition of having P, one should accept that any Q has P. I think that no one would deny the validity of this criterion (see also Williams 1986, 60).

Furthermore, Korsgaard's earlier worries about arguments from the first person do not seem to apply to the argument from the sufficiency of agency. In *The Sources of Normativity*, Korsgaard's main worry about first-personal arguments appears to be that they are *prudential*

arguments: I have to regard my humanity as the source of *my* values; you have to regard your humanity as the source of yours. But this leaves open the question of why I should value *your* humanity. Understood this way, "[T]he individual is thought to be self-interested, or, at least, self-interest is taken to be an uncontroversial source of rational norms" and "Rational justifications of morality must then show that self-interest gives the individual some reason to participate in a moral system (Korsgaard 1996, 132).[5] If this, indeed, is what the first personal argument says, then Korsgaard is right to deem it wildly implausible and hopeless for justifying interpersonal morality.[6]

But notice that Korsgaard's more recent argument from the sufficiency of agency that compares it to citizenship is clearly different from the first-personal argument she criticizes in the passage above. Korsgaard's recent argument from the sufficiency of agency is clearly not a prudential argument. Instead, the argument is that I necessarily have to value my own humanity, my rational nature, simply by virtue of being an agent. Given that this is a characteristic I share with other agents, I must also value their rational nature. In short, the difference is between thinking that agents have a purely prudential reason for valuing their own humanity, and thinking that agents necessary ought to value their own humanity simply by virtue of understanding themselves as an agent. The latter 'ought' is not (merely) a prudential ought but first and foremost a transcendental ought (cf. Beyleveld and Bos 2009).

The important point for present purposes is that if Korsgaard's argument from the sufficiency of agency succeeds, there would be no longer a need to rely on the controversial Wittgensteinan publicity of reasons argument to justify interpersonal morality. Or at least, pace Korsgaard (1996), there seems to be no principled reason why this argument could not lead to conclusions of interpersonal morality.

The argument from the sufficiency of agency, as mentioned earlier, is the second step of Korsgaard's master argument to ground interpersonal morality. The first step, recall, is that any agent has to value her own humanity. Therefore, to provide a justification for interpersonal morality, i.e. for Korsgaard's master argument to be successful, we need not only the argument from the sufficiency of agency to be valid taken on itself, but it should deliver the desired conclusions when combined with the first step of Korsgaard's argument. Given that the argument from the sufficiency of agency taken on itself seems to be uncontroversial, I will thus evaluate whether this argument *together with* the first step of Korsgaard's master argument (which is supposed to show that your own humanity has value) indeed leads to a categorical and universal principle of interpersonal morality.[7] I will argue that even if we assume that the argument from the sufficiency of agency is valid, the argument from the sufficiency of agency, together with what I take to be the strongest reconstruction of her argument for the value of your own humanity, does not lead to moral obligations to others.

[5] This is Korsgaard's description of the project of Hobbes and Gauthier. But she claims that "some neo-Kantian justifications [among which Gewirth's] proceed, or anyway might be thought to proceed, in a similar way (Korsgaard 1996, 132–33).

[6] I don't think, however, that Korsgaard's interpretation is charitable to those who have actually made the sufficiency of agency argument. Korsgaard's (1996) description of arguments from the first person is for instance uncharitable towards Gewirth's (1978) argument (cf. Beyleveld 2015).

[7] The importance of evaluating the argument as a whole and not only the universalization step of the argument is noted by several critics of Gewirth's argument from the sufficiency of agency (McMahon 1986; Kramer and Simmonds 1996; Chitty 2008). For replies to these critics in the context of Gewirth's argument see Beyleveld and Bos (2009) and Beyleveld (2013).

3 The Argument for the Value of Your Own Humanity

There are different interpretations of Korsgaard's argument for the value of your own humanity. The first, and most dominant interpretation is often referred to as the 'regress of identities argument' (referred to from now on as the 'regress argument'). The regress argument states that we can and should always ask 'why-questions' about our actions and also about our reasons for action ('Why do I want to drink coffee?' 'Because I want to be more focused.' 'Why do I want to be more focused?' 'Because I want to finish this paper.' 'Why do I want to finish this paper?' etc.). Korsgaard claims that this regress of why-questions cannot be brought to an end by one of our particular 'practical identities' - "description[s] 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 1996, 101) – but only by the value of humanity.

The regress argument is widely criticized (Kerstein 2001; Sussman 2003; Timmermann 2006; Stern 2011; Street 2012). Recently, Robert Stern, for instance, objects to Korsgaard's claim that we should always doubt our particular practical identities, by suggesting that certain particular practical identities could also stop the regress of why-questions (Stern 2011, 87–88). Because for certain people their particular identities can stop the regress, only hyper-reflective agents, agents who in fact always doubt their particular practical identities, have to value their own humanity (because for others valuing their particular practical identities might be sufficient).[8] If this criticism is correct, applying the argument from the sufficiency of agency to the conclusion of the regress argument obviously does not lead to any categorical and universal principles of interpersonal morality. For in this case the regress argument does not show that I have to value humanity simply by virtue of being an agent, but only by being a hyper-reflective agent. The most the argument could show is that insofar as I am a hyper-reflective agent I do not just have to value my own humanity, but also the humanity of all other hyper-reflective agents.

The alleged failure of the regress argument is often taken as an indication that Korsgaard fails to justify the value of (your own) humanity. Stern has, however, put forward a reconstruction and defence of a different possible interpretation of Korsgaard's argument, which I will refer to as 'the source of reasons argument.'[9] In the remainder of this paper, I will focus on this argument. The reason for this is that it seems to be the

[8] I think that the main problem with the regress of reasons argument is that when Korsgaard discusses the contingency of our practical identities, she shifts between claiming that we *may* always doubt our practical identities identity (Korsgaard 1996, 122) to the much stronger claim that we *should* do so and that we therefore need a reason to adopt a specific practical identity (Korsgaard 1996, 258).

[9] A third candidate is Korsgaard's reconstruction of Kant's argument for the value of humanity in her paper 'Kant's Formula of Humanity' (1986). In this argument, the value of humanity is understood to be a necessary condition of the possibility of acting for an objectively good end (Korsgaard 1986, 190). Korsgaard writes that "if humanity is not regarded and treated as unconditionally good then nothing else can be objectively good" (Korsgaard 1986, 198). This argument thus presupposes that there is such a thing as a categorical imperative. This means that the argument is conditional: if there is a categorical imperative, then one must value humanity. Korsgaard describes this argument and the relation between the categorical imperative, the objective good and the value of humanity as follows: "having established that if there is a categorical imperative there must be something that is unconditionally valuable, Kant proceeds to argue that it must be humanity" (Korsgaard 1986, 194). Because the argument is contingent on already accepting the categorical imperative, I will not discuss this argument further and instead will focus on two more ambitious arguments, which do not assume the categorical imperative.

strongest possible reconstruction of Korsgaard's argument for the value of your own humanity. Stern, for instance, writes: "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 argument can be made to work" (Stern 2011, 74).[10]

Stern begins by taking his cue from Korsgaard's description of the argument for the value of humanity which she finds in Kant according to which "humanity, as the source of all reasons and values, must be valued for its own sake" (Korsgaard 1996, 122). Stern takes the central idea of this argument to be the following: "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" (Stern 2011, 90). The idea is thus that if one acts, one must think that the end of one's action is good. If one subsequently rejects the idea that the end of one's action is good in itself (i.e. if one rejects moral realism), the end of one's action can only be good because it is good for oneself. But if the end of one's action is only good if it is good for oneself, one must regard oneself, i.e. the life of someone who acts for reasons, as valuable because otherwise one's end cannot be good.

Why do you have to see yourself as valuable in order to act for a reason? Stern puts forward the following reconstruction of Korsgaard's argument along the general lines of Korsgaard's interpretation of Kant's argument for the value of humanity.

1. To rationally choose to φ, you must take it that φ-ing is the rational thing to do.
2. Since X [i.e. the object of your action] 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 (Stern 2011, 90).

The distinctive element of Stern's reconstruction of Korsgaard's argument has to do with premise 4: the idea that practical identities are not valuable as such. According to Stern, Korsgaard does not necessarily have to rely on something like the regress argument, i.e. the

[10] I do not discuss the exegetical question of whether Stern's reconstruction of Korsgaard's argument is plausible in light of the original text. Instead, I discuss Stern's reconstruction of Korsgaard's argument as the strongest possible Korsgaardian argument for the value of humanity.

idea that we should always question our particular practical identities, to reach the conclusions that particular practical identities are not valuable as such. Instead, Stern states that

> to see value in any particular identity as such is to be committed to realism, to thinking that being a father, an Englishman, a university lecture 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 do to. But, as we have seen, Korsgaard also takes such realist positions to be problematic, so can perhaps use such arguments [i.e. the argument that practical identities are not valuable as such] here, without appealing to the regress considerations at all (2011, 91–92).

So, according to Stern, Korsgaard can justify the idea that a particular practical identity is not valuable as such by relying on her rejection of moral realism. Practical identities could only be valuable as such if one commits oneself to realism. If one rejects realism, particular practical identities cannot be valuable as such. Assuming for the sake of argument that moral realism should indeed be rejected,[11] practical identities cannot be valuable as such.

Although practical identities are not valuable as such, we need a practical identity in order to have a reason for action (premise 2), so there must be another source for the value of our particular practical identities (premise 3). According to Stern's reconstruction of Korsgaard, if one needs a practical identity to act on reasons, and if particular practical identities are not valuable as such, the only value of particular practical identities is that they make it possible for an agent to act on the basis of reasons (the second part of step 4): "such identities have the general capacity of enabling the agent to live a life containing reasons: because I have whatever particular practical identities ... I can then find things to be valuable and act rationally accordingly, in a way that gives me unity as a subject" (Stern 2011, 92).

The argument from step 5 to step 8 of the argument subsequently proceeds as follows:

> 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 is to value your humanity (Stern 2011, 92).

This is a very condensed argument, but roughly the idea is that if practical identities can only be valuable insofar as they make it possible to live a life containing reasons, you must think that it is important that your life has a rational structure, and this is subsequently taken to mean that you must value your rational nature (step 7) and your humanity (step 8).

I think that the main conclusion of the argument is reached in step 5: that you must think that your having a life containing reasons and values is important for your practical identities to be valuable. Since Stern, following Korsgaard herself, does not further define 'rational nature'

[11] Stern disagrees with this assumption (2012) and it is a recurring objection to Korsgaard that she has not provided a convincing objection to moral realism (see e.g. Regan 2002, 272; Darwall 2006, 231). I think that in general a Kantian constructivist is well advised to try to provide an argument for interpersonal morality which does not presuppose the failure of moral realism. For the sake of argument, however, I will assume that Korsgaard is right to assume the failure of moral realism and subsequently evaluate the argument on its own terms.

and 'humanity' except by reference to each other and to the idea of leading a life which contains reasons, I understand the remaining steps of the argument (6–8) as saying roughly the same thing in different words. My understanding is therefore that the concepts of 'rational nature' and 'humanity' do not mean anything over and above living a life containing reasons, i.e. leading a rationally structured life (I come back to this below, and I also come back to the question of what it means to 'value' your own 'humanity' within the terms of this argument).

In what follows, I will assume that Stern's reconstruction of Korsgaard's argument is valid. Subsequently I raise the question whether the source of reasons argument, when combined with the argument from the sufficiency of agency, can justify moral obligations towards others.

4 Why Korsgaard Cannot Justify Moral Obligations Towards Oneself

My objection, in a nutshell, is this: even if the argument for the value of humanity, as Stern reconstructs it, is valid, it is unclear whether the argument could lead to any *normative* conclusions, because it is unclear how acknowledging the value of humanity could translate into any normative requirements for action. This implies that even if the conclusion of this argument can be successfully universalized through the argument from the sufficiency of agency, so that one has to value the humanity of all other agents as well, it would fail to justify any normative principle of interpersonal morality.

This objection might sound counterintuitive at first. After all, is it not obvious that one could do all kinds of things to express respect for the value of one's own and someone else's humanity? Helping people in need, for instance, seems a clear case of expressing respect for the value of other people's humanity. And is there not also a plurality of ways in which one could fail to express this kind of respect? Torturing other people is clearly inhumane. The same goes for slavery, living in severe poverty, human trafficking and so on. In other words, is it not obvious that the value of humanity translates into certain normative requirements?

I agree that on our *common-sense* understanding of the 'value' of 'humanity', activities like torture or slavery very obviously fail to express respect for the value of humanity. Korsgaard appeals to this common sense understanding of the value of humanity when, in the *The Sources of Normativity,* she writes that her argument shows "that Enlightenment morality is true" (Korsgaard 1996, 123) and that "to value yourself just as a human being is to have moral identity, as the Enlightenment understood it" (Korsgaard 1996, 121). Although Korsgaard does not elaborate on what exactly she means by 'Enlightenment morality', I assume it refers, among other things, to the idea of human rights, which protect people from some of instances of inhumane treatment mentioned above and much more. In her recent work on the value of humanity she likewise writes:

> "[Valuing humanity] involves respecting the rational choices of other people, and making ourselves fit for the normative standing it confers on us, by developing and preserving our rational powers. And it also involves, quite simply, caring about ourselves and each other, not only as rational but as natural beings, whose interests we declare, through our moral legislation, to be worthy of realization, promotion, and pursuit. (Korsgaard forthcoming, 37)

The question, however, is whether Korsgaard's actual argument for the value of humanity, discussed in the previous section, justifies *this* particular understanding of the value of

humanity. After all, what it means to value one's humanity, on Kosgaard's argument, depends on what exactly is contained in the premises of the argument. This means we can only assess the success of the argument for the value of humanity if we are clear about what it means to 'value' one's (own) 'humanity' in terms of Korsgaard's argument and not in terms of our common sense understanding of these terms. So the question is: what does it mean, exactly, to say that one must value one's (own) humanity in the context of Stern's reconstruction of Korsgaard's source of reason argument? I will discuss this question in two steps. First, by analysing what if means to value your *own* humanity. Second, by analysing what it means to value the humanity of (all) *other* persons.

First, what does it mean to value one's own humanity? On Stern's reconstruction of Korsgaard's argument, 'humanity' refers to 'rational nature', where rational nature is understood as a life containing reasons and values. Or as Korsgaard writes "we are self-conscious rational animals, capable of reflection about what we ought to believe and to do" (Korsgaard 1996, 46). Neither Korsgaard nor Stern define humanity except by reference to this idea of leading a rationally structured life.

In addition, if one looks closely at the argument, humanity could not possibly mean anything more than 'one's rational nature'. Practical identities provide reasons, and practical identities can only be valuable to the extent to which one thinks it is valuable that they provide reasons. That is, practical identities are important insofar as they contribute to a rationally structured life, which means that they can only be important insofar as one thinks that leading a rationally structured life is important. On this definition of humanity, the *absence* of humanity is thus not an inhumane life, as we would normally perhaps think of it, but means living a life without any reasons and values. Perhaps such a life would be the life of certain non-human animals, or robots or inanimate things. In other words, the absence of humanity means that one would be a mere object and not an agent; one would be "the mere undergoer of ... experiences" (Korsgaard 1989, 120). Humanity is thus another word for 'rational agency' or an 'agential life'. This should make us cautious when it comes to claims about 'respecting' or 'valuing' our humanity. These will, after all, not be conclusions about, for instance, basic human rights but rather claims about not leading a robot-life or valuing our leading a life containing reasons.

Having discussed what we should, on Korsgaard's account, take 'humanity' to mean, we can now turn to the question: what does it mean to *value* (one's own) humanity and how does this value translate into norms of action? Stern interchangeably speaks about the need "to value your humanity" (Stern 2011, 92) and about thinking that "having a life containing reasons and values is *important*" (Stern 2011, 90 my emphasis). In addition, he claims that denying the value of humanity means thinking that "you and your life were utterly worthless, pointless, meaningless—that in your eyes, you were valueless" (Stern 2011, 89).

Korsgaard proposes to understand the value of humanity as referring to a normative status or standing and she claims that norms should subsequently be understood as expressing respect for this standing or status (Korsgaard 1996, 145, forthcoming, 25–32). She writes that "duties are expressions of respect, not for the property of rationality, but for the legislative standing that it confers upon us" (Korsgaard forthcoming, 31–32). To value one's own humanity, then, is to think or see that there is a value to having a life containing reasons and values and that this gives you a normative status. Duties subsequently express respect for the status one has by virtue of being a rational agent.

However, if this is what it means to value one's humanity, it is unclear how this value could generate any normative constraints on action. For the options seem to be the following: either one is an agent and therefore one necessarily has to think that it is important to have a life containing reasons – important in the sense that otherwise one could not have a reason for action. Alternatively, one does not think that this is important, but in that case one would not be an agent in the first place because, according to Korsgaard's argument, one necessarily has to value one's humanity if one is an agent.

In other words, if valuing your humanity (on Korsgaard's technical definition of humanity) is a necessary condition of the possibility of acting for a reason, there can be, according to this argument, no agents who do not value their humanity but who *should* value their humanity. There is no reason to value one's humanity if one is not an agent (plants have no 'reason' to value their humanity), but if one is an agent one necessarily has to think that having a life containing reasons is important because otherwise one could not have a reason in the first place (again, assuming that the argument is valid).

The same point could be made in a slightly different way. Recall that the starting point of Korsgaard's argument is that we inescapably act for reasons, i.e. that we inescapably understand ourselves as agents. The conclusion of the argument is that we should 'value' our 'humanity', i.e. that we should value having a life containing reasons and values. However, for the value of humanity to lead to any normative conclusions there should be ways in which we could fail to understand ourselves as agents. As Korsgaard herself acknowledges, "[T]here is no normativity if you cannot be wrong" (Korsgaard 1996, 161). But if one cannot fail to understand oneself as an agent (because, according to Korsgaard, one *inescapably* understands oneself as an agent), one could not fail to value one's humanity either, given that valuing one's humanity is a necessary condition of the possibility of agency. And if one does not understand oneself as an agent (e.g. robots, rocks and plants), one does not have a reason to value one's own humanity. The value of humanity, in other words, does not provide a source of (moral) normativity, because, according to this argument, understanding oneself as an agent and valuing one's humanity cannot be separated.

In her discussion of suicide in *The Sources of Normativity*, Korsgaard seems to admit this point. Here she acknowledges that someone who commits suicide cannot be said to act immorally, i.e. cannot be said to fail to value his or her humanity: "It is hard to say of one who commits such suicide that he has done wrong, for he has violated no value in which he still believes" (Korsgaard 1996, 162–63). Towards the end of the same section she generalizes from the discussion of suicide, by suggesting that it is not an option for us to fail to lead a rational life and subsequently to fail to value our humanity. She writes: "there really is a sense in which, being human, and as long as we go on living, we *have to* engage in rational action. Animal action, unreflective action, is not open to us" (Korsgaard 1996, 164).

The point here is not that Korsgaard should be able to say that suicide is in conflict with valuing your own humanity, but that her remarks about suicide seem to generalize to other potential duties towards oneself. Consider, for instance, the question of whether selling yourself into slavery would be against valuing your own humanity. Following Korsgaard's discussion of suicide, even selling yourself into slavery would be compatible with valuing your own humanity, because as, as she writes, as long as we go on living we have to engage in rational action (and we therefore necessarily value our humanity).

If we necessarily have to engage in rational action, and if we must value our humanity if we are to engage in rational action, valuing our own humanity cannot lead to any normative

requirements. Treating oneself inhumanely is, on Korsgaard's argument for the value of humanity, simply not possible for us.

5 Why Korsgaard Cannot Justify Moral Obligations Towards Others

But what about the interpersonal case? Could one not fail to value the humanity of other persons? If so, the value of humanity could lead to normative constraints in the interpersonal domain, even if it does not lead to any intrapersonal normative claims. That is, assume, for the sake of argument, that the argument from the sufficiency of agency succeeds and that one must not only value one's own humanity but also the humanity of all other agents, i.e. that one must value that others live a life containing reasons. What are the normative implications of this conclusion?

Let me start by saying that, in a trivial sense, valuing the humanity of other agents does lead to a normative requirement. After all, one could fail to *think* that others have value without ceasing to be an agent oneself. Valuing the agency of others is not something that I must do in order to be able to act on reasons myself, but, following the argument from the sufficiency of agency, something I must accept on pain of consistency, i.e. on pain of denying that my humanity is valuable by virtue of being an agent. Valuing the humanity of other agents is thus something one must do even if one currently does not. This normative requirement is, however, trivial as long as it remains unclear whether acknowledging the value of agency actually generates any normative requirements on actions, i.e. whether it tells us how (not) to treat other agents.

So this leads to the following question: what would it mean to express respect for the status that others have by virtue of leading a life containing reasons? Again, just like in the intrapersonal case, there seems to be little, if nothing, one could do to undermine this status. After all, Korsgaard claims that leading a life containing reasons is *inescapable* not just for me but for any agent. One can still be an agent on Korsgaard's definition while being injured, coerced or threatened, or even when being imprisoned or enslaved. The reason for this is that even in the most deprived situations it seems to be at least *possible* to "back up and bring [an] impulse into view" (Korsgaard 1996, 93).

It is actually quite difficult to think of any action which would qualify as failing to express respect for another person's humanity, if respecting a person's humanity means nothing more than respecting the fact someone should be able to be a rational agent. Even killing another person does not necessarily seem to be incompatible with valuing the humanity of another person. After all, the moment the other person is still alive he or she might, on Korsgaard's view, still qualify as an agent. Whereas the moment the other person is dead, there is no longer a reason to value his or her agency in the first place (because he or she is no longer an agent). Killing another person would only be problematic if the argument requires respect for a *continued* existence of the agent. But, at least as it stands, this is not what the argument argues for, and it is unclear how it could follow from the premises in Stern's reconstruction of Korsgaard's argument. Valuing your (and subsequently another person's) continued existence as an agent does not seem to be necessary for having a reason for action in the same way as valuing living a rationally structures life is necessary.

The problem here thus seems to be very similar to the intrapersonal case: if agency is inescapable (not just for you but also for others) then it is hard to see how one could fail to act in a way that respects the value of humanity of other persons. After all, what is needed to be an

agent, according to the argument for the value of humanity, is merely to have the capacity to make a rational choice. It is hard to violate this requirement because it seems that one cannot act in such a way as to undermine the capacity of rational choice, without at the same time undermining the very basis for assigning a normative status to this person, i.e. the presence of this capacity. So again, it does not seem to be possible to treat another person wrongfully because either this person remains an agent and therefore the capacity which deserves respect is still intact (which means one did not act immorally), or the person is no longer an agent in which case there is no longer a reason to value its humanity, because it does no longer fulfil the sufficient condition for its humanity to have value.

Korsgaard, however, seems to be assume that agency can be preserved or developed by yourself and by others (Korsgaard forthcoming, 37). This might be true on certain conceptions of agency. For instance, conceptions that involve having a rational plan of life. The conception of agency on which Korsgaard's argument relies, however, is a much more minimal conception of agency, which mainly entails the capacity for reflection and self-consciousness. This conception of agency is not one which can be preserved or developed but one we, as Korsgaard herself explicitly stresses, inescapable have. It is the inescapability of Korsgaard's conception of agency, which makes that it is hard, if not impossible, to act in such a way as to fail to express respect for this capacity.

What I hope to have shown is that the underlying problem with Korsgaard's argument is that what it means to value humanity depends both on how one understands value and humanity and on how one reconstructs Korsgaard's argument for the value of humanity. The main problem with Korsgaard's argument is that 'humanity' is introduced as a technical term and should not be confused with the everyday understanding of this concept. I am afraid that it is only if we lose sight of the crucial distinction between the concept of humanity in Korsgaard's argument and the everyday understanding of this concept that the value of humanity has the moral–political depth that Korsgaard wants it to have, and that we can conclude that justifying the value of humanity shows that "Enlightenment morality is true" (Korsgaard 1996, 123). At best, then, Korsgaard's argument for the value of humanity as articulating the source of (interpersonal) moral *normativity* is incomplete; at worst, the argument fails because the idea of the value of humanity is too trivial to generate any norms of action.

6 Concluding Remarks

In this paper, I have reconstructed and critically analysed an overlooked argument by Korsgaard for interpersonal morality: the argument from the sufficiency of agency. Although this argument is less controversial than the widely discussed argument from the publicity of reasons, I have argued that Korsgaard's argument from the sufficiency of agency cannot lead to universal and categorical obligations to others. This failure, I should stress again, has less to do with the argument from the sufficiency of agency itself, and more to do with the details of Korsgaard's argument for the value of humanity on which the success of the argument from the sufficiency of agency crucially depends. The failure of Korsgaard's argument leaves open the possibly that other Kantian constructivist arguments which rely on the argument from the sufficiency of agency, such as Gewirth (1978), are more successful.

It lies beyond the scope of the paper to defend an alternative Kantian constructivist argument for interpersonal morality, but let me end this paper by briefly mentioning two

reasons why a Gewirthian argument from the sufficiency of agency might be more successful than Korsgaard's argument from the sufficiency of agency. First, Gewirth does not use the language of values, but the language of rights and duties in his argument from the sufficiency of agency. That is, Gewirth focuses on the *normative* structure of agency from the very start. Second, Gewirth does not try to show that we must value our *humanity*, but that every agent has a right to what he calls the 'generic features of agency'. These generic features of agency refer to the means that are necessary to engage in any action whatsoever. That is, Gewirth does not argue for the moral protection of our agency or humanity as such, but for the moral protection of the means that are necessary to engage in any action whatsoever. This is important because one can fail to take or protect these means in the same way as one can be instrumentally irrational (Beyleveld and Bos 2009).

Although much more needs to be said in defence of Gewirth's argument for interpersonal morality, it should be stressed that given that the problems for Korsgaard's argument from the sufficiency of agency where caused by her reliance on certain ideas of 'value' and 'humanity', an analogous argument which does not rely on these notions will at least not face the same problems as Korsgaard's argument. Whether or not Gewirth's argument faces other objections, is a topic for another paper.

Let me conclude by saying that given the well-known problems of the publicity of reasons argument, the argument from the sufficiency of agency might be the best hope for Kantian constructivists to justify interpersonal morality, even if Korsgaard's version of the argument does not work.

Acknowledgements I would like to thank Rutger Claassen, Marcus Düwell, Fleur Jongepier and Ingrid Robeyns for their helpful comments on various earlier drafts of this paper.

References

Apel K-O (1980) The a priori of the communication community and the foundations of ethics: the problem of a rational foundation of ethics in the scientific age. In: Towards a transformation of philosophy. Marquette University Press, Milwaukee, pp 225–300

Beyleveld D (2013) Williams' false dilemma: how to give categorically binding impartial reasons to real agents. J Moral Philos 10(2):204–226

Beyleveld D (2015) Korsgaard v. Gewirth on universalization: why Gewirthians are Kantians and Kantians ought to be Gewirthians. J Moral Philos 12(5):573–597

Beyleveld D, Bos G (2009) The foundational role of the principle of instrumental reason in Gewirth's argument for the principle of generic consistency: a response to Andrew chitty. King's Law J 20(1):1–20

Chitty A (2008) Protagonist and subject in Gewirth's argument for human rights. King's Law J 19(1):1–26

Darwall S (2006) The second-person standpoint: morality, respect, and accountability. Harvard University Press, Cambridge

De Maagt S (2018) It Only Takes Two to Tango: against Grounding Morality in Interaction. Philos Stud. https://doi.org/10.1007/s11098-018-1150-3

Gert J (2002) Korsgaard's private-reasons argument. Philos Phenomenol Res 64(2):303–324

Gewirth A (1978) Reason and morality. University of Chicago Press, Chicago

Habermas J (1990) Discourse ethics: notes on a program of philosophical justification. In: Moral consciousness and communicative action. MIT Press, Cambridge, pp 43–115

Kerstein SJ (2001) Korsgaard's Kantian arguments for the value of humanity. Can J Philos 31(1):23–52
Korsgaard CM (1986) Kant's formula of humanity. Kant-Studien 77(2):183–202
Korsgaard CM (1989) Personal identity and the unity of agency: a Kantian response to Parfit. Philos Public Aff 18(2):101–132
Korsgaard CM (1996) The sources of normativity. Cambridge University Press, Cambridge
Korsgaard CM (2007) Autonomy and the second person within: a commentary on Stephen Darwall's the second-person standpoint. Ethics 118(1):8–23
Korsgaard CM (2009) Self-constitution: agency, identity, and integrity. Oxford University Press, Oxford
Korsgaard CM (forthcoming) Valuing our humanity. In: *Respect for persons*, edited by Oliver Sensen and Richard Dean. Oxford University Press, Oxford
Kramer MH, Simmonds NE (1996) Reason without reasons: a critique of Alan Gewirth's moral philosophy. South J Philos 34(3):301–315
LeBar M (2001) Korsgaard, Wittgenstein, and the Mafioso. South J Philos 39(2):261–271
McMahon C (1986) Gewirth's justification of morality. Philos Stud 50(2):261–281
O'Neill O (1986) The public use of reason. Political Theory 14(4):523–551
Regan DH (2002) The value of rational nature. Ethics 112(2):267–291
Skidmore J (2002) Skepticism about practical reason: transcendental arguments and their limits. Philos Stud 109(2):121–141
Skorupski J (1998) Rescuing moral obligation. Eur J Philos 6(3):335–355. https://doi.org/10.1111/1468-0378.00065
Stern R (2011) The value of humanity: reflections on Korsgaard's transcendental argument. In: Smith J, Sullivan P (eds) Transcendental philosophy and naturalism. Oxford University Press, Oxford, pp 74–95
Stern R (2012) Constructivism and the argument from autonomy. In: Lenman J, Shemmer Y (eds) Constructivism in practical philosophy. Oxford University Press, Oxford, pp 119–137
Street S (2012) Coming to terms with contingency : Humean constructivism about practical reason. In: Lenman J, Shemmer Y (eds) Constructivism in practical philosophy. Oxford University Press, Oxford, pp 40–59
Sussman D (2003) The authority of humanity. Ethics 113(2):350–366
Timmermann J (2006) Value without regress: Kant's 'formula of humanity' revisited. Eur J Philos 14(1):69–93. https://doi.org/10.1111/j.1468-0378.2006.00244.x
Wallace RJ (2009) The publicity of reasons. Philos Perspect 23(1):471–497
Williams B (1986) Ethics and the limits of philosophy. Harvard University Press, Cambridge