Philos Stud (2019) 176:2767–2783
https://doi.org/10.1007/s11098-018-1150-3

It only takes two to tango: against grounding morality in interaction

Sem de Maagt[1]

Published online: 3 August 2018

Abstract Most Kantian constructivists try to ground universal duties of interpersonal morality in certain interactions between individuals, such as communication, argumentation, shared action or the second-person standpoint. The goal of this paper is to present these, which I refer to as arguments from the second-person perspective, with a dilemma: either the specific kind of interaction that is taken as a starting point of these arguments is inescapable, but in that case the argument does not justify a universal principle of interpersonal morality. Or interaction does have a principle of interpersonal morality among its necessary conditions of possibility, but such forms of interactions are merely optional. I argue that proponents of arguments from the second-person perspective have failed to provide a convincing response to this dilemma and that this failure is systematic. This suggests that the success of Kantian constructivism depends on the success of arguments from the first person.

Keywords Kantian constructivism · Transcendental arguments · Morality · First person · Second person

1 Introduction

Kantian constructivism is an ambitious ethical project which aims to justify a categorical and universal principle of interpersonal morality while avoiding some of the common metaphysical, epistemological and motivational problems of (robust) moral realism. It tries to achieve this goal by employing a transcendental argument which aims to show that the acceptance of a categorical and universal principle of

✉ 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

interpersonal morality is necessary insofar as one understands oneself as engaging in action or interaction[1]—where, given that (inter)action is somehow inescapable for A (i.e. you, me or any other human being), it logically follows that A must accept this moral principle. For instance, Christine Korsgaard (1996) has tried to show that we inescapably understand ourselves as an agent and that anyone who understands herself as an agent necessarily has to accept the value of humanity, insofar as valuing one's humanity is a necessary condition for the possibility of agency. Importantly, Korsgaard argues that agents should not just necessarily value their own humanity, but also the humanity of all others agents.[2]

Recently, several critics have argued that Kantian constructivism cannot justify any categorical reasons for action, let alone (interpersonal) moral reasons, because there is no conception of agency or interaction that is inescapable in the relevant sense (Enoch 2006, 2011; Silverstein 2015), or because even if agency is inescapable it cannot lead to any normative conclusions (Street 2012). Although I think that these objections can be answered (De Maagt forthcoming), I want to bracket these well-known objections in this paper. For even if Kantian constructivists are able to justify some categorical reasons for action this would still be a far cry from justifying any categorical and universal principle having to do with the question of how we ought to treat others—the latter being the main ambition of Kantian constructivism. If Kantian constructivism cannot justify universal interpersonal morality it would be a significantly less interesting ethical position to discuss in the first place. Therefore, I shall focus on the often-overlooked question of how Kantian constructivism tries to justify claims of universal interpersonal morality specifically, e.g. on how it tries to show that someone should not only necessarily value her own humanity but also the humanity of all other persons (Korsgaard 1996, p. 130).[3]

What I will refer to as transcendental arguments from the first-person perspective (in short: arguments from the first person) hold that it is possible to argue from the claim that A must accept whatever are the necessary preconditions of understanding herself as an agent (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). What I will refer to as transcendental arguments from the second-person perspective (in short: arguments from the second person), on the other hand, claim that it is impossible to argue from the necessary preconditions of individual agency to interpersonal morality. Instead, arguments from the second person propose to take as their starting

[1] I will come back to this distinction between action and interaction below. For the sake of simplicity, I use action and agency interchangeably.

[2] With a 'universal principle of interpersonal morality' I thus refer to a principle which entails moral respect for all persons (collective universality). In another sense of the term, the claim that every agent has to value their own humanity is also a universal duty because this is something that every agent has to accept (distributive universality).

[3] One reason why the question of interpersonal morality is often not discussed in recent meta-ethical debates is that if Kantian constructivism cannot justify any categorical normative claim it can certainly not justify any categorical claim of interpersonal morality (Street 2012, p. 48f15).

point A's involvement in some form of interaction, such as communication, argumentation, shared action or the second-person standpoint, and subsequently explore the necessary conditions of the possibility of understanding oneself as being engaged in these forms of interaction. By focusing on the necessary conditions of the possibility of interaction versus the necessary conditions of the possibility of action, the idea is that there no need to cross the bridge between the personal and the interpersonal in the first place, because the necessary conditions of the possibility of interactions are interpersonal from the very start. Or at least this is what transcendental arguments from the second person aim to show.

Most recent attempts at putting forward a transcendental argument for interpersonal morality are arguments from the second person (Apel 1980; Habermas 1990; O'Neill 1986; Darwall 2006; and to some extent Korsgaard 1996, 2009). The main goal of the paper is to argue that despite the existence of important differences between these arguments, they are all susceptible to the same dilemma: either the specific kind of interaction that is taken as a starting point of these arguments is inescapable, but in that case it is unconvincing that the necessary conditions of the possibility of these inescapable forms of interaction include the acceptance of a *universal* principle of interpersonal morality, i.e. a principle of morality which entails moral respect for all other persons. Alternatively, some form of (universal) interaction does have a universal principle of interpersonal morality among its necessary conditions of possibility, but in that case the said interaction seems to be merely optional and the moral principles are therefore not categorically valid.[4] This suggests that the viability of transcendental arguments in ethics (at least to the extent to which they aim to justify categorical and universal principles of interpersonal morality) depends on the success of transcendental arguments from the first person.

The structure of the paper is as follows: in Sect. 2 I will elaborate on the distinction between transcendental arguments from the first person and transcendental arguments from the second person and I will briefly reconstruct the motivation behind the Kantian constructivist move towards the second person. In Sect. 3, I present the dilemma for transcendental arguments from the second person. Finally, in Sect. 4 I will discuss some responses to this dilemma and argue that these responses presuppose what they have to argue for, namely, that all individuals with certain (second-personal) capacities deserve equal moral respect.

[4] This dilemma could be understood as an instantiation of a more general potential dilemma for Kantian constructivism and transcendental arguments: either the starting point of a transcendental argument is inescapable, but in that case no interesting (moral) conclusions can follow from the starting point, or Kantian constructivism might succeed in justifying interesting (moral) conclusions, but in that case the starting point turns out to be merely optional (not inescapable). For similar (instantiations of this) potential dilemma(s) see e.g. Tiffany (2012).

2 First person versus second person

Transcendental arguments from the first person hold that A must not only accept whatever are the necessary preconditions of understanding herself as an agent (e.g. value A's humanity), but that A is also required to assign the same status to the necessary preconditions of the agency of other individuals (e.g. value the humanity of all other agents). This argument typically proceeds in two steps. The first step tries to show that A must accept whatever are the necessary conditions for the possibility of her own agency insofar as she understands herself as an agent:

1) Agency is inescapable for A
2) Y is a necessary condition of the possibility of agency
3) Therefore, A must Y (e.g. 'value his/her humanity')

The next step involves trying to universalize the above conclusion in order to include a moral concern for the necessary conditions for all other agents:

4) Agency is a sufficient condition for accepting Y
5) Therefore, A must not only Y (e.g. value his/her humanity') but also U (e.g. value the humanity of all persons)
(where U refers to the necessary preconditions of the agency of all persons)[5]

The basic idea here is that if agency is a sufficient condition for one's humanity having value (or whatever the necessary preconditions of agency are) one necessarily has to accept that not only one's own humanity but the humanity of all other agents has value (or whatever the necessary preconditions of agency are). Alan Gewirth (1978, 104–112) therefore refers to this kind of argument as the 'argument from the sufficiency of agency.'

However, most of the recent transcendental arguments for interpersonal morality that have been proposed in ethics start not from A's self-understanding as an agent, but from A's involvement in some form of interaction, such as communication, argumentation or shared action (see introduction for references). Very roughly, what I will call transcendental arguments from the second person claim that A is inescapably understands herself as being involved in some form of interaction and that, insofar as interaction has certain necessary conditions of possibility, A has to accept whatever are the necessary conditions of the possibility of this kind of interaction. Crucially, the claim is that the necessary conditions of the possibility of interaction entail the acceptance of a categorical and universal principle of interpersonal morality. In contrast to an argument from the first person, an argument from the second person thus argues for claims about interpersonal morality directly, instead of using a two-step argument. An argument from the second person roughly has the following structure:

[5] An argument with this structure can be found in Gewirth (1978, 104–12) and some of the work of Korsgaard (1986, 196; forthcoming, 28–31). In Korsgaard (1996, 2009) she is, however, critical of transcendental arguments from the first person and she puts forward a transcendental argument from the second person (I will come back to this below)..

1) Interaction is inescapable for A
2) Y is a necessary condition of the possibility of interaction
3) Therefore, A must Y
(where Y refers to the acceptance of a categorical and universal principle of interpersonal morality)

Both arguments from the first person and from the second person start from the internal perspective of A acting or interacting. That is, these arguments do not (merely) try to provide a conceptual analysis of action or interaction, but they try to analyse what A is necessarily committed to by understanding herself as an agent or as someone who engages in certain forms of interaction. The necessary conditions of action and/or interaction should be accepted by A because these are what makes it possible for A to understand herself as acting or interacting. The internal, first-personal or second-personal perspective is crucial to understand how Kantian constructivists try to derive normative conclusions from agency or interaction.

The popularity of arguments from the second person is motivated by recurring worries about arguments from the first person. Some argue against arguments from the first person by claiming that the reason why A has to accept the necessary conditions of the possibility of agency is purely prudential and that it is not similarly prudential to extend moral concern to the necessary conditions of possibility of someone else's agency (MacIntyre 1985, 2nd:80; Williams 1986, 61; Habermas 1990, 101; Korsgaard 1996, 134). Others claim that 'agency' is not a sufficient condition for accepting Y (premise 4), because is not agency as such but A's particular agency that is a sufficient condition for accepting Y (Chitty 2008). And again others claim that an argument from the first person simply provides the 'wrong kind of reason' for morality (Darwall 2006, 16).

The point here is not to evaluate these objections but merely to highlight the main idea behind these kinds of objections in order to understand the motivation behind transcendental arguments from the second person.[6] The idea that motivates these arguments is that there seems to be an unbridgeable gap between the personal and the interpersonal, so that one cannot get from categorical self-referring reasons to categorical other-referring reasons. An argument from the second person does not have to bridge this gap because it begins from A's involvement in some form of interaction and, if successful, its conclusions are therefore interpersonal from the very start.

There are important differences between different arguments from the second person, for instance regarding the precise starting point of their argument (their conception of interaction) and in how they try argue from their preferred conception of interaction to morality. O'Neill (1986) and Apel (1980), for instance, put forward an extension of Wittgenstein's private language argument through an investigation of the normative conditions of the possibility of communication and

[6] The main problem with these objections is that they presuppose that arguments from the first person are purely prudential. The reason for accepting whatever are the necessary preconditions of understanding oneself as an agent are, however, not only prudential but they are transcendental, i.e. they are commitments that every agent necessarily has to accept by pain on contradicting that one understands oneself as an agent. See also Beyleveld and Bos (2009).

argumentation.[7] Korsgaard (1996, 2009) puts forward an analogy to Wittgenstein's private language argument by trying to show that not only language but also reasons are inherently public by exploring the necessary conditions for the possibility of shared action.[8] Darwall (2006) on the other hand adopts a broadly Stawsonian-Austinian approach which tries to articulate the so-called 'felicity conditions' of the second-person standpoint.[9]

Nevertheless, I think that, despite these differences, it is possible to detect a common core to these arguments. First, they believe that there is a form of interaction that is inescapable for A and that interpersonal morality should somehow be grounded in this form of interaction. Second, they hold that participating in this form of interaction has a normative dimension, i.e. that by engaging in some form of interaction A is committed to certain 'rules of the game' (e.g. moral respect for the other). Finally, and most importantly, they hold that these 'rules' have a universal scope, i.e. they entail moral respect for *all* others.[10]

In what follows, I will formulate a dilemma for arguments from the second person. Although I cannot go into the details of all the specific arguments within the scope of this paper, I will try to show that despite the differences mentioned above they all face the same dilemma, that they have failed to provide a convincing response to this dilemma, and that this failure is systematic. Although this analysis will inevitably lack some nuance because of the lack of in depth discussion of the different positions, I hope that this lack of nuance will be compensated for by what it shows about the general prospects and problems of the Kantian constructivists project.

3 A dilemma for arguments from the second person

The first horn of the dilemma states that there might be certain forms of interaction that are indeed inescapable, but that it is unconvincing that the necessary conditions of the possibility of these inescapable forms of interaction include the acceptance of a *universal* principle of interpersonal morality.[11] That is, even if one accepts that

[7] Apel, for instance, presents his project as a "transcendental-philosophical radicalization of the later Wittgenstein's work" (Apel 1980, 269).

[8] "I meant […] to be making an argument for the publicity of reason that is analogous to Wittgenstein's argument for the publicity of meaning. Wittgenstein's argument, as I understand it, is intended to show that meaning can't be normative at all—you can't be wrong—unless it is public. My argument was meant to show that reasons cannot be normative at all unless they are public" (Korsgaard 2009, 196f12).

[9] A further difference is that they all put forward slightly different moral principles and/or slightly different interpretations of the categorical imperative. But this difference should not concern us here.

[10] Another reason to group these authors together under one label is that nearly all of these authors crucially rely on Wittgensteinian insights concerning the publicity of language. One might argue that Darwall is an exception to this rule. However, Darwall relies on Strawson's notion of reactive attitudes, which could be interpreted as a Wittgensteinian theory. I will not, however, pursue this suggestion here.

[11] Illies (2003, 74) puts forward this argument against Karl-Otto Apel's argument from the second-person. Similar objections have been put forward against the publicity of reasons argument of Korsgaard (1996) by LeBar (2001), Gert (2002), Skidmore (2002) and Wallace (2009). Below I briefly illustrate the

some form of interaction is inescapable, and that engaging in interaction commits us to certain norms, it is unclear why this would commit us to a *universal* moral principle. If A systematically excludes some others from interaction or fails to express moral respect for certain persons why should interaction as such be impossible for A?

Let me try to illustrate this objection by briefly discussing Korsgaard's (2009) argument from the second person in some more detail. Before looking at the argument itself, it is important to stress that Korsgaard's (2009) argument does not nicely fit into the distinction I made between transcendental arguments from the first person and transcendental arguments from the second person. On the one hand, Korsgaard tries to show that any agent necessarily has to value her own humanity by using a transcendental argument with a first-personal structure (1996, 123, 2009, 23). On the other hand, she (1996, 2009) uses a transcendental argument from the second person to argue for the thesis that all reasons are public, 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). Korsgaard combines the conclusion of this 'argument from the publicity of reasons' with the conclusion of her argument for the value of your own humanity to conclude that if one necessarily has to value one's humanity, and if reasons are necessarily public, so that the normative force of a reason is shared with others, then every agent necessarily has to value the humanity both in him- or herself and in all other agents. 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).

The reason for focussing on Korsgaard in this section is that however exactly we understand the relation between the two steps of her argument, her argument for interpersonal morality has a distinctively second-personal structure. In addition, Korsgaard (1996, 133) explicitly criticizes arguments from the first person at least insofar as they are supposed to lead to conclusions about interpersonal morality, including Gewirth's argument from the sufficiency of agency. Finally, Korsgaard is one, if not the most influential Kantian constructivists. In what follows, I concentrate on her most recent defence (2009) of the argument from the publicity of reasons.

Korsgaard puts forward the following argument by example:

> Suppose you and I are related as student and teacher, and we are trying to schedule an appointment. "Stop by my office right after class," I say, thinking that that will be convenient for me, and hoping that it will also be convenient for you. It isn't, as it turns out. "I can't," you say, "I have another class right away." So I have to make another proposal. It's important to see why I do

Footnote 11 continued
first horn of the dilemma by applying it to Korsgaard's recent (2009) defence of the argument from the publicity of reasons.

> have to do this: it's because having the meeting is something that we are going to do together (Korsgaard 2009, 192).

Korsgaard's point is that the student's reason necessarily also has to be a reason for Korsgaard at the same time, because they are planning to perform a shared action, i.e. they are planning to have a meeting *together*. According to this argument, the publicity of reasons, in the Korsgaardian sense of your reason being normative for me, is thus a necessary precondition of the possibility of acting together. Thus, Korsgaard concludes, "to perform a shared action, each of us has to adopt the other's reasons as her own, that is, as normative considerations with a bearing on her own case" (Korsgaard 2009, 192). This argument could be roughly summarized as follows:

1) Shared action is inescapable for A
2) The publicity of reasons is a necessary condition of the possibility of shared action
3) Therefore, A must accept the publicity of reasons

However, even if we accept that engaging in this kind of shared action commits one to treating the other's reasons as normative for you, it remains unclear why one should treat the reasons of *all others* as normative for you, and not just the reasons of the persons one plans to perform a shared action with. Although Korsgaard puts forward this transcendental argument as an argument for a claim about the very nature of reasons, it is unclear how the publicity of reasons in general (and not just in the cases in which one actually engages in shared action) is supposed to be a necessary condition for the possibility of engaging in specific instances of shared action, such as the shared action between teacher and student. In other words, it is unclear how Korsgaard tries to argue from the necessary conditions for the possibility of shared action with a less than universal audience (e.g. the shared action of teacher and student), to a conclusion about reasons in general and ultimately to a conclusion about universal interpersonal morality.

In order to argue for the publicity of reasons, Korsgaard invites the reader to consider the alternative, in which she does *not* treat the reasons of the student as her own reasons. Korsgaard suggests that, in this case, she might respond to the student's remark that she cannot make it because she has another class by saying, "[W]ell, just skip it" (Korsgaard 2009, 193).

According to Korsgaard, there is something deeply problematic about this alternative, for "obviously, we can't relate at all on those terms. So if that's how it is, no personal interaction is going to be possible" (Korsgaard 2009, 193). In this case, the only kind of interaction that is possible is purely strategic interaction with the student in which her reasons are treated merely as "possible tools and obstacles, things that might help me to achieve my ends or get in my way" (Korsgaard 2009, 193). But this, Korsgaard claims, would amount to "a kind of war, or combat" (Korsgaard 2009, 194).

But even if is true that living in a constant state of war is not a real option for us, this argument is not sufficient to justify the categorical and universal moral

obligations that Korsgaard wants to argue for. The problem with Korsgaard's argument is that even if we grant that we cannot live in a constant state of war and that engaging in shared action commits us to treating the reasons of the other person as our own reasons, this does not commit us to engaging in shared action with *all* other human beings and consequently to valuing the humanity of *all* other agents. What, on Korsgaard's account, would be wrong with the racist who treats some other people's reasons as his or her own reasons, but not the reasons of someone of another race? The racist refuses to engage in shared action with specific others and is therefore in these cases not committed to any of the necessary presuppositions of shared action. His or her life can hardly be described as a constant war.

Note that Korsgaard has not even shown that the racist is committed to valuing the humanity of a person of another race at the moment they are interacting, because there seem to be many ways of interacting (e.g. market transactions) which do not qualify as shared action on Korsgaard's definition of shared action. Thus, even if we assume for the sake of argument that shared action commits one to the value of humanity of the persons one is engaging in shared action with, Korsgaard has not shown that the same applies to other forms of interaction which are not quite shared actions. The only moment the racist is committed to valuing the humanity of a person of another race would be when he or she engages in shared action with this person. Although this is not an insignificant conclusion, it is still far removed from the Kantian ethics that Korsgaard aims to justify.

Korsgaard thus presents the reader with a false dilemma. Korsgaard suggests that one has to value the humanity of all agents or one's life is a constant war, but this overlooks the possibility that one might only engage in shared action with certain particular others or only value the humanity of some others.

Other proponents of arguments from the second person present the reader with similar false dilemmas. According to Apel, the choice is either "paranoid-autistic loss of self" (1975, 268) or one has to respect the universal language community. According to O'Neill, one is either "left in solitary and thoughtless silence" (1986, 539) or one has to accept the categorical imperative. As should have become clear by now these dilemmas neglect the fact that there might be all kinds of ways of arguing, communicating and reasoning that do not require (moral respect for) a universalist audience. I might argue, communicate or reason with person X, and satisfy the relevant second-personal conditions, without being obligated by virtue of the necessary conditions for the possibility of interaction to have to engage in this way with person Y. In other words, it is unconvincing that argumentation, communication, reasoning, shared action or second-person interaction completely *break down* if one does not accept a universal principle of morality.

At one point Korsgaard writes that "it takes two to make a reason" (1996, 138). I think that Korsgaard might be right about this. She is wrong however to suggest that it requires anything *more* than two to make a reason. It might take two to make a reason, but it does not seem to require the 'Kingdom of Ends.' What this shows is that the strength of transcendental arguments from the second person is at the same time its weakness. The strength of a transcendental argument from the second person is that it starts from an interpersonal perspective, which means that there is no need to bridge the gap between the personal and the interpersonal. Its weakness,

however, is that the gap between the personal and the interpersonal is simply replaced by different gaps: between reasons that should at least be communicable to others and reasons that all other agents should adopt as a principle of action; between recognizing the dignity of the participants in a particular second-person relation and recognizing the dignity of all agents.

The first horn of the dilemma presupposes that one engages in interactions with a less than universal audience (e.g. the interaction between teacher and student). This raises the worry that these kinds of interactions do not commit one to a universal moral principle. Things, however, seem to be different in cases of interactions with a universal audience. After all, these kinds of interactions might indeed have a commitment to a universal principle among its necessary conditions of possibility. In other words, in order to escape the first horn of the dilemma one might try to focus on universal interactions from the very start, instead to trying to get to moral universality through the analysis of less than universal interactions (first horn of the dilemma).

The second horn of the dilemma, however, states that although some forms of (universal) interaction might indeed have a universal principle of interpersonal morality among its necessary conditions of possibility, in that case interaction seems to be optional and the moral principles are therefore not categorically valid.

Consider, for instance, Karl-Otto Apel who suggests that the kinds of claims that Wittgenstein makes about *all* languages and *all* language games have a universal audience and subsequently commit the speaker to morally respect all persons (Apel 1975, 259). Apel's idea is thus that there are certain speech acts that necessarily presuppose a universal audience. The examples he uses are claims about language. Another example might be discussions about morality (cf. Habermas 1990). The idea is thus that there are certain forms of interactions with a universal audience which commit one to a universal moral principle.

However, even if one assumes that a universal language game is a necessary condition of the possibility of making *universal* claims about language or of arguing about morality, this only shows that one is necessarily committed to a universal language community insofar as one makes these kinds of claims (Illies 2003, cf. 86). But surely making these kinds of claims about language is only an optional practice; writing books like *Philosophical Investigations* can hardly be called inescapable. The same goes for discussions about morality. The idea that there are certain universal interactions which have the acceptance of a universal principle among their necessary conditions of possibility is therefore insufficient to prove that anyone who engages in argumentation is necessarily committed to a universal language community (including the universal moral norm which, according to Apel and others, is among the necessary conditions of the possibility of this universal language community).

Unsurprisingly most arguments from the second person do not try to show that a universal moral principle is a necessary condition for the possibility of interacting with all others, but that a commitment to a universal principle follows from interacting as such, including interaction with some specific others (such as the example of the teacher and the student). In the remainder of the paper I will therefore focus on the ways in which proponents of the second-personal strategy have tried to respond to the first horn of the dilemma.

4 The implicit assumption: respect for all persons

Proponents of arguments from the second person have responded to, or anticipated, this dilemma in different ways. In this section I briefly consider some of these replies and argue that they ultimately presuppose, rather than argue for, the claim that we have to morally respect all persons with the appropriate second-personal capacities. I will do so by discussing Darwall's recent anticipation and response to this kind of objection in some detail and subsequently indicate how similar (unsuccessful) responses are put forward by others as well. The reason for focussing primarily on Darwall's argument is that he has provided one of the most recent and most exhaustive replies to the objection discussed in the previous section.

Darwall illustrates his analyses of the normative dimensions of the second-person perspective through the following example: suppose someone stands on your foot and you either demand or request that this person remove her foot from yours. In that case, Darwall states that "you and she commonly presuppose that she *can* freely comply if she finds your request or demand one she could not reasonably reject, regardless of what she desires or how strongly she desires it" (Darwall 2006, 245). In other words, if you address someone, you have to presuppose that the addressee is capable of responding to reasons. In addition, Darwall claims that this second-personal address not only presupposes the *capacity* (what Darwall calls 'second-personal competence') to respond to this address as a person but also presupposes the *normative authority* to engage in this relation as a free and equal person: "[the second-person standpoint] commits both parties … to the one normative standing that is inescapable for them both, that of a free and rational person or will as such" (Darwall 2006, 246). Darwall's point is that if you make a request of someone, you have to presuppose that you have the authority to make this request and that the addressee can acknowledge your authority, and consider your request or demand, as a free and rational person.

Up until this point, Darwall's argument is very similar to Korsgaard's argument from shared action. Both start from an interaction with a specific other, and subsequently point out that this interaction commits one to certain normative conclusions (the sharing of reasons or acknowledging the freedom and equality of the other). Darwall, however, puts forward an additional argument to argue for a universal moral conclusion. Darwall's argument is the following:

> Any second-personal authority at all can exist only if it can be rationally accepted by free and rational agents as such. But for that to be true there must be grounds for such an acceptance, and whatever interests free and rational agents have as such would have to be among such grounds. It is conceptually necessary, moreover, that free and rational agents have an interest in not being subject to others' arbitrary will since that would, by definition, interfere with the exercise of their free and rational agency. As this interest must be among the grounds that free and rational agents have for accepting any authoritative demands at all, it necessarily supports a demand, as free and rational, against being subject to demands that cannot be so justified. It follows, again, that second-personally competent agents have the authority to demand that they

> not be subject to mere impositions of will, that is, to demanding (coercive) conduct that cannot be justified second-personally (Darwall 2006, 274).

It is a complex argument, and, sadly, Darwall says very little to explain the different steps of the argument. But it seems to consist of the following steps:

1) Accepting the second-personal authority of, for instance, the person who stands on your foot means that you should conceive of the person who stands on your foot not just as that particular person, but (also) as a free and rational agent, i.e. as someone who can respond to reasons.[12]
2) This can only be true if there are interests that a free rational agent has simply insofar as he or she is a free and rational agent.
3) One of these interests is to not be subject to another person's arbitrary will, and this, according to Darwall, supports a (legitimate) demand to not be subject to the will of another person.
4) Anyone with second-personal competence has the authority to demand that they will not be subject to another person's arbitrary will, i.e. that anyone with second-personal competence has dignity.

Darwall thus tries to show that having second-person capacities or competencies is sufficient to have equal dignity. He repeats this point in more recent writing:

> It is sufficient to be a moral agent in this sense [i.e. as an equal member of the moral community] that a being has the *psychic capacities* necessary to enter into relations of mutual accountability, that is, to take a second-person perspective on himself and others and regulate his conduct from this point of view (Darwall 2010, 221–222 my emphasis).[13]

Thus, one is not only committed to respecting the persons with whom one actually engages in second-personal interaction but also any other person who has the competences or capacities to engage in this type of interaction. Darwall concludes that

> whenever we take up the second-person standpoint at all and address any claim or demand whatsoever, we are committed to the shared second-personal authority of *any* second-personally competent being (moral agent or person) on which the ideas of moral obligation and rights depend (Darwall 2010, 224 my emphasis).

Darwall thus holds that we have to respect the equal dignity of any person with second-personal competencies or capacities, i.e. any person who is able to have reactive attitudes.

The problem with this argument, however, is that I fail to see how this conclusion follows from an analysis of the second-person standpoint. The reason for this is that there is an unexplained gap between the following:

[12] Darwall writes that "the addressee is… conceived of as a-person-who-happens-to-stand-in-that-normative-relation" (Darwall 2006, 268).

[13] Cf. "second-personal reasons are no less valid for an agent who happens to reject them but who has the *capacity* to take up a second-person perspective and accept them" (Darwall 2007, 59 my emphasis).

1) The claim that insofar as one engages in a second-person relationship one has to assume that the person one stands in such a relation to is a free and rational agent.
2) The claim that having second-person competencies is *sufficient* for deserving moral respect.

Even if it is true that, as Darwall suggests, there are certain interests that persons have simply by virtue of being a free and rational agent, this does not yet show that I have to respect these interests independent of my being involved in any second-person relationship with these persons. For instance, even if I accept that the person who stands on my foot has an interest in not being subject to my will, at least insofar as I engage in a second-personal relation with this person, it does not follow that all persons with the same competences can make the same (legitimate) demand, nor that I consequently have to respect their dignity, because it is not necessary that I actually have a second-personal relation with them. It does not necessarily follow from the claim that I have to respect the interest of the person who stands on my foot *as* an interest of her as a rational and free agent that I also have to respect the same interest of all other persons with the same competences, unless it is these competences themselves that are morally deserving of respect. But Darwall cannot simply help himself to the latter assumption; the justificatory work has to come from second person and not from a pre-given commitment to the equal dignity of any person with certain capacities.

One might object to this by claiming that it is exactly Darwall's goal to show that a commitment to the equal dignity of *any* person with certain capacities and not just the equal dignity of the person one actually engages in a second-personal interaction that is part of the necessary conditions for the possibility of second-personal address. I fail to see, however, why specific second-personal interactions would only be possible if I accept the equal dignity of any person with certain capacities (versus the specific person one actually interacts with). Or at least, as it stands Darwall fails to argue show that this is truly a necessary condition for the possibility of second-personal address. In other words, it is unclear why second-personal competency is a sufficient condition of deserving moral respect. Or at least I fail to see how this conclusion is supposed to follow from Darwall's analysis of the second-person standpoint.

Other proponents of arguments from the second person make similar assumptions about the equal moral status of all persons with the relevant second-personal capacities. Apel, for instance, writes that "all beings who are *capable* of linguistic communication must be recognized as persons" (Apel 1980, 259 my emphasis). Apel claims that these persons should be respected because "they are potential participants in a discussion, and the unlimited justification of thought cannot dispense with any participant, nor with any of his potential contributions to a discussion" (Apel 1980, 259). But again, one might wonder why the fact that someone is a *potential* participant in a discussion justifies the idea that respecting this person is a necessary precondition of the possibility of argumentation per se. Why, in other words, would argumentation become impossible if I do not respect all *potential* participants in a discussion? Similar worries apply to O'Neill's reliance on

assumptions about "other's capacities to act, to suffer and to be influenced" (2000, 193–94) in determining the (universal) scope of morality. Again the problem is that O'Neill does not argue for the claim that anyone with certain capacities has an equal moral standing, nor does she explain why this principle is among the necessary conditions of the possibility of communication. In other words, O'Neill presupposes that every being with certain capacities has an equal moral standing. And again it is hard to see how this principle could be justified in terms of the necessary conditions of the possibility of communication.[14]

Apel, O'Neill and Darwall thus ultimately refer to agential or second-personal capacities to justify the claim that all human beings, or at least all human beings with these capacities, have equal moral standing. It is, however, unclear how this appeal could work within the structure of an argument from the second person. The problem is that it is unclear how a principle which states that all individuals with agential capacities have equal moral status could follow from an analysis of the necessary conditions of the possibility of argumentation, communication or second-personal interaction so that it follows that A is necessarily committed to this principle insofar as he or she engages in certain forms of interaction. Instead, this principle seems to be simply presupposed instead of argued for.[15]

In other words, what is doing the relevant normative work to justify a universal principle of interpersonal morality is not the analysis of the necessary conditions of the possibility of some form of interaction but the idea that any individual with certain capacities deserves moral respect.

[14] In a recent text, O'Neill relies on a slightly different argument. She writes "Why, one might ask, does ethical reasoning have to offer reasons *to all*? Many have thought otherwise, and have claimed that ethical reasoning may, even must, draw on the established beliefs, conventions or ideologies of particular communities and traditions, so will be fit to reach some but not others. Kant saw this as a mistake. Reasoning that is premised on local or contingent beliefs, conventions or traditions is incomplete reasoning because it offers reasons only to some and not to all, and assumes various contingent beliefs as premises. Such reasoning, as Kant sees it, is inevitably less than fully public. It is *heteronomous reasoning*, in that it assumes some other source or authority, and while many admirable acts reflect heteronomous reasoning, the touchstone of reason is *not* to rely on extraneous or arbitrary assumptions. Fully reasoned belief and action seek to rely on reasons that could be offered to *all* others, so aim to be fully public in the Kantian sense (O'Neill 2013, 223; a similar argument can be found in Ronzoni and Valentini 2008). Although this needs more argumentation than I can provide here (due to reasons of space), I think that this argument does not succeed because O'Neill has not given us any reason to believe that relying on a restricted community of communication is the same as relying on a non-existing pre-established authority. Say that in deliberation I exclude those persons who only reason on the basis of selfish premises. Why would I, in that case, invoke a pre-established authority (cf. Barry 2013, 25)? I do not necessarily have to presuppose the truth of my ideology or religion in order to have good reasons to exclude certain others from deliberation. More importantly, O'Neill has not given us any reason to believe that communicating with a *universal* community of communication is any less arbitrary than communicating with a *restricted* community of communication. Why is it more arbitrary to only reason with some, instead of reasoning with all? This judgement seems to rely on a moral judgement that we *should* reason with all others. I do not see how this kind of universalism could follow simply from what it means to reason.

[15] Not surprisingly, some critics of the authors discussed above therefore claim that these arguments can only lead to categorical and universal principles of interpersonal morality by relying on an independent, realist value of certain (agential) capacities (see e.g. Langton 2007; Barry 2013; Gilabert 2005a, b, 2006).

5 Concluding remarks

I have argued that transcendental arguments from the second person, despite their differences, are all susceptible to the same dilemma: they cannot show that there is an inescapable form of interaction which at the same time has a universal moral principle among its necessary conditions of possibility. This shows that, at best, current transcendental arguments from the second person are incomplete and, at worst, that these arguments simply fail to justify categorical and universal principles of categorical morality. If my argument is correct, the viability of the Kantian constructivist project hinges on the plausibility of transcendental arguments from the first person.[16]

It lies beyond the scope of the paper to defend an argument from the first person. But let me end this paper by briefly mentioning some reasons why I think that an argument from the first person might succeed where arguments from the second person fail. Arguments from the first person try to justify universal, interpersonal morality by analysing what an agent is necessarily committed to by understanding herself as an agent. Given that an agent has to claim a certain moral status for herself (e.g. value her own humanity), simply by virtue of understanding herself as an agent, it is inconsistent to deny the same moral status to other agents. After all, if agency is a sufficient condition for having a certain moral status, one must accept that any agent has this moral status. So if it is true that one has to accept that one has a certain moral status simply by virtue of understanding oneself as an agent, any agent has to accept that all agents have this moral status. Such an argument from the first person can therefore show that any individual with certain capacities deserves moral respect. But whether or not such an argument can actually work, is a topic for a different paper.

Acknowledgements An earlier version of this paper was presented at the 'Moral Justification and Transcendental Argumentation' workshop at Utrecht University (2017). I am grateful for the comments and questions I received on this occasion. Special thanks to Rutger Claassen, Marcus Düwell, Katrin Flikschuh, Fleur Jongepier and Ingrid Robeyns for their helpful comments on various earlier drafts of this paper. I would also like to thank an anonymous reviewer for his or her constructive comments and suggestions.

References

Apel, K.-O. (1975). The problem of philosophical fundamental-grounding in light of a transcendental pragmatic of language. *Man and World, 8*(3), 239–275.

[16] To my mind the strongest defence of a transcendental argument from the first person is put forward by Gewirth (1978, 104–12). For a defence of Gewirth's argument against critics see Beyleveld and Bos (2009) and Beyleveld (2013, 2015).

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* (pp. 225–300). Milwaukee: Marquette University Press.

Barry, M. (2013). Constructivist practical reasoning and objectivity. In D. Archard, M. Deveaux, N. Manson, & D. Weinstock (Eds.), *Reading Onora O'Neill* (pp. 17–36). Abingdon: Routledge.

Beyleveld, D. (2013). Williams' false dilemma: How to give categorically binding impartial reasons to real agents. *Journal of Moral Philosophy, 10*(2), 204–226.

Beyleveld, D. (2015). Korsgaard v. Gewirth on universalization: Why Gewirthians are Kantians and Kantians ought to be Gewirthians. *Journal of Moral Philosophy, 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 Journal, 20*(1), 1–20.

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

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

Darwall, S. (2007). Reply to Korsgaard, Wallace, and Watson. *Ethics, 118*(1), 52–69. https://doi.org/10.1086/528725.

Darwall, S. (2010). Precis: The second-person standpoint. *Philosophy and Phenomenological Research, 81*(1), 216–228.

De Maagt, S. (forthcoming). *Why Humean constructivsts should become Kantian constructivists*. Unpublished Manuscript.

Enoch, D. (2006). Agency, Shmagency: Why normativity won't come from what is constitutive of action. *Philosophical Review, 115*(2), 169–198.

Enoch, D. (2011). Shmagency revisited. In M. S. Brady (Ed.), *New waves in metaethics*. London: Palgrave Macmillan.

Gert, J. (2002). Korsgaard's private-reasons argument. *Philosophy and Phenomenological Research, 64*(2), 303–324.

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

Gilabert, P. (2005a). A substantivist construal of discourse ethics. *International Journal of Philosophical Studies, 13*(3), 405–437.

Gilabert, P. (2005b). The substantive dimension of deliberative practical rationality. *Philosophy and Social Criticism, 31*(2), 185–210.

Gilabert, P. (2006). Considerations on the notion of moral validity in the moral theories of Kant and Habermas. *Kant-Studien, 97*(2), 210–227.

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

Illies, C. (2003). *The grounds of ethical judgement: New transcendental arguments in moral philosophy*. Oxford: Oxford University Press.

Korsgaard, C. M. (1986). Kant's formula of humanity. *Kant-Studien, 77*(2), 183–202.

Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge: Cambridge University Press.

Korsgaard, C. M. (2007). Autonomy and the second person within: A commentary on Stephen Darwall's the second-person standpoint. *Ethics, 118*(1), 8–23.

Korsgaard, C. M. (2009). *Self-constitution: Agency, identity, and integrity*. Oxford: Oxford University Press.

Korsgaard, C. M. (forthcoming). Valuing our humanity. In O. Sensen & R. Dean (Eds.), *Respect for persons*. Oxford: Oxford University Press.

Langton, R. (2007). Objective and unconditioned value. *Philosophical Review, 116*(2), 157–185.

LeBar, M. (2001). Korsgaard, Wittgenstein, and the Mafioso. *Southern Journal of Philosophy, 39*(2), 261–271.

MacIntyre, A. (1985). *After virtue*. London: Duckworth.

O'Neill, O. (1986). The public use of reason. *Political Theory, 14*(4), 523–551.

O'Neill, O. (2000). Distant strangers, moral standing and porous boundaries. In *Bounds of justice* (pp. 186–202). Cambridge: Cambridge University Press.

O'Neill, O. (2013). Responses. In D. Archard, M. Deveaux, N. Manson, & D. Weinstock (Eds.), *Reading Onora O'Neill* (pp. 219–243). Abingdon: Routledge.

Ronzoni, M., & Valentini, L. (2008). On the meta-ethical status of constructivism: Reflections on G.A. Cohen's 'facts and principles'. *Politics, Philosophy & Economics, 7*(4), 403–422.

Silverstein, M. (2015). The Shmagency question. *Philosophical Studies, 172*(5), 1127–1142.
Skidmore, J. (2002). Skepticism about practical reason: transcendental arguments and their limits. *Philosophical Studies, 109*(2), 121–141.
Street, S. (2012). Coming to terms with contingency: Humean constructivism about practical reason. In J. Lenman & Yonatan Shemmer (Eds.), *Constructivism in practical philosophy* (pp. 40–59). Oxford: Oxford University Press.
Tiffany, E. (2012). Why be an agent? *Australasian Journal of Philosophy, 90*(2), 223–233.
Wallace, R. J. (2009). The publicity of reasons. *Philosophical Perspectives, 23*(1), 471–497.
Williams, B. (1986). *Ethics and the limits of philosophy*. Cambridge, MA: Harvard University Press.