Michihito YOSHIME (Osaka University)

On the Precedence of the First-personal Point of View in Contemporary Kantian Moral Arguments

Introduction

Christine M. Korsgaard once noted:

> Although it is universally acknowledged that human beings are in fact social animals, modern moral philosophers have usually not considered it allowable to help themselves to this fact in arguments aimed at justifying morality. … But our social nature is deep, in the sense that it is the nature of our reasons that they are public and shareable, then justifications of morality can and should appeal to it (Korsgaard [1996] 2014, 135).

This diagnosis seems correct, and is even more applicable when it comes to the work of contemporary Kantian moral philosophers from John B. Rawls to Korsgaard herself. The core of Rawls's theory of justice consists in taking the 'original position' behind the 'veil of ignorance,' where each individual should represent a free and equal citizen and choose his/her act without bias or favor. Korsgaard, who accepts Ludwig Wittgenstein's refutation of the idea of private language, argues that we can exchange and share the reasons for our acts, despite these reasons themselves coming from the 'reflective endorsement' of our individual 'practical identities.'

Discourse ethics, which was founded and developed by Jürgen Habermas and Karl-Otto Apel, tries, on the other hand, to reconstruct Kant's moral philosophy in terms of a theory of intersubjective discourse. According to this theory, our ethical norms must be 'discursively redempted' through critical trials of our 'validity claims' of the norms. The 'principle of universalization' then functions as the criterion for assessments. We can safely say that this theory appeals to our social relations to justify morality, because the point of this theory is that discursive redemption, and discourse itself, is a shared action: no one is able to perform it or represent it by him/herself. However, this theory has also been attacked from a seemingly Kantian standpoint that rejects the idea of morality based on consensus, and calls for a return to the convictions of independent individuals.

In this short paper, I would like to give a critique of moral arguments that still presuppose a first-personal starting point, in accordance with the above diagnosis by Korsgaard. Needless to say, problematizing subjectivity is a well-known story and covers broad fields in

philosophy. However, I am neither going to argue this point in a wider context, which would include perspectives such as language-analytical, phenomenological, psychoanalytic or mind-body arguments, nor will I provide any detailed reconstruction of Rawls's, Korsgaard's or even Kant's moral philosophy. Here, I will just point out that the method of advancing moral arguments that presuppose a first-personal starting point does not really seem Kantian. Rather, it runs contrary to Kant's potential theory of the others, which can be derived from his doctrine of transcendental idealism. Such a modest conclusion would not be trivial in that the aforementioned Kantian moral philosophers would aim to proclaim themselves as Kantian henceforth.

For this purpose, I will pick up a seemingly Kantian criticism of discourse ethics made by Albrecht Wellmer, who was Habermas's former assistant in Frankfurt. I am not going to do this with regard to the problem of the distinction between morality and law, which is better known as Wellmer's critique of Habermas, but with regard to criticism of the rationality of consensus itself. I will investigate the logical consequence of a first-personal reading of Kant by means of the contrast between transcendental and empirical idealism, and try to show that discourse ethics is in better accordance with Kant's original thought. The argument will proceed as follows: (1) I will investigate Wellmer's criticism of Habermas from the first-personal point of view and cast doubt on the precedence of that point of view. (2) I will try to reveal that there is a certain assumption about *the other* behind the act of giving precedence to the first-personal point of view. (3) I will contrast that order of precedence with Kantian thought on *the other* and point out that first-personalist theories seem totally contrary to Kant, despite their self-attribution as Kantian.

1

As is well known, Habermas introduced the universal principle (U), which every valid norm has to fulfill, as the 'moral principle,' and the principle of discourse ethics (D), which is derived from (U) as follows:

> (U) *All* affected can accept the consequences and the side effects its *general* observance can be anticipated to have for the satisfaction of *everyone's* interests (and these consequences are preferred to those of known alternative possibilities for regulation).
>
> (D) Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a practical discourse* (Habermas 1990, 65-66).

Habermas considers this principle to represent impartiality, which Kant attempted to capture via the categorical imperative. Strictly speaking, the compatibility of these principles with Habermas's universal pragmatics, developed in the 1970s, is unclear because the role of the counterfactual anticipation of an 'ideal speech situation', once placed as a regulative principle of each factual communication, is obscured. Yet the central idea of this communicative or discursive theory of reason is that morality lies not in our first-personal consciousness but in the intersubjective consensus, which we should reach through moral discourse. Habermas argues that one cannot make his/her maxim valid alone, in advance of the moral discourse in which the maxim should be assessed. At this point, Habermas's view differs from those contemporary Kantians who start their argument from the first-personal perspective. And this is the very point that Albrecht Wellmer questions, in his understanding, from the viewpoint of Kant.

Roughly sketched, Wellmer argues that we had better understand Kant's categorical imperative as prohibiting non-generalizable maxims (cf. Wellmer 1991, 123-126), and, furthermore:

> If we understand the categorical imperative in this way, the 'monologic' character of Kant's moral principle is not such a serious problem as it appears to Apel and Habermas. For if *I* am unable to will that a way of acting should become a universal rule, then *we* cannot will it either (otherwise *I* should be able to do so as well). We might equally express the point like this: in moral judgement I am above all confronted with myself (Ibid. 153).

Although Weller admits that Kant was mistaken in simply assuming that my 'being able to will' must necessarily coincide with that of all other rational beings, he does not fully accept Habermas's discourse-theoretical reconstruction of Kantian morality. Because, according to him, the concept of a consensus achieved on the basis of reasons presupposes the concept of a personal conviction. If I have reasons for agreeing, then this means, precisely, that I consider a validity claim to be true. But, so he argues, then the truth does not follow from the rationality of the consensus, but from the appropriateness of the reasons by which I was convinced of its truth. "It is only from the point of view of those involved in the situation that consensual rationality appears to be identical with truth" (Ibid. 161). And, in so far as a moral discourse is involved here, this objection cannot be dismissed as irrelevant for discussing morality but truth[1]. Wellmer could give precedence to our first-personal conviction about the normative validity of our maxims over the consensus of moral discourses. This need not be the central part of his criticism of discourse ethics. On the contrary, he suggests the concept

[1] I have criticized Wellmer before with regard to his hermeneutic truth theory (cf. Yoshime 2016, 270-282).

of the ethics of dialogue, in which the 'principle of dialogue' does not replace Kant's moral principle but works as the main one among many supplementary principles for the moral principle. So, he does not actually intend to sustain the claim of the first-personal standpoint. Nonetheless, I consider his argument on conviction important for understanding a tendency in contemporary "Kantian" standpoints as a whole.

Let us begin with a simple but not trivial question. While arguing that convictions precede consensus, he has also admitted that one's being able to will cannot be regarded as identical to that of all other rational beings. And if Habermas is right in saying that the categorical imperative has an impartial mind in it then, apparently, a valid maxim is such that all other rational beings can also will it, i.e., be convinced of its validity. Then, how is it at all justifiable to give precedence to a first-personal conviction over an agreed or 'exchanged and shared' (in Korsgaard's sense) conviction of all members of a moral discourse? Wellmer emphasizes the fallibility of each consensus achieved in a moral discourse, and that must be one of the reasons why he suggests making discourse ethics into a weaker ethics of dialog. But it does not amount to a reason for us to regard the first personal conviction as infallible, or at least as more reliable.

His answer to this question turns out to be less Kantian. That is, he renounces the universal validity of the categorical imperative and understands it as saying, simply, something like, "Act according to your normative convictions" or "Do what you (believe that you) ought to do" (Ibid. 207). With this 'minimal' interpretation of the categorical imperative, his previous interpretation of it in terms of prohibition of what one cannot will, of course, loses its significance. It was only temporarily employed by Wellmer to make first-personal convictions also look somehow valid in the context of universal morality. And we can and should ask here, again: Why is first-personal conviction given precedence over consensus? Even if such a conviction 'contains no trivial requirement' as he hopes, and even if we admitted that consensus in general is fallible, we could still discuss the matter to find new viewpoints, new evidence, or new arguments for or against his conviction. Is there any good reason to believe that an individual agent is able to make better judgements about his/her actions than two or more jointly discussing agents? Wellmer's thought on our convictions strongly suggests the superiority of the first-personal to intersubjective relations with others.

2

The only plausible reason for persistence in the first-personal point of view is that the proponents of this view understand consensus as simply agreeing with others, and think

that following such agreements is *heteronomous* (in Kant's sense)[2]. As mentioned above, Wellmer himself does not maintain such a view. However, it is likely that there are some Kantian moral philosophers who do reason in this way. Even for Korsgaard, who accepts Wittgenstein's argument against private language and discusses our social nature, it is decisively important that *I* make the 'reflective endorsement' on my reason for acting. She states:

> The reflective structure of human consciousness requires that you identify yourself with some law or principle which will govern your choices. It requires you to be a law to yourself. And that is the source of normativity. So, the argument [on our 'practical identity'] shows just what Kant said it did: that our autonomy is the source of obligation (Korsgaard [1996] 2014, 103-104).

So, having my reason discussed in a moral discourse and following its agreed conclusion is heteronomous, as long as the conclusion is not in accordance with my practical identity. This could be morally problematic in cases such as that of an egoist who wills to act autonomously. On the other hand, Korsgaard has to make her argument consistent with the fact that we can exchange and share our reasons. She does this by characterizing the difference between one's relationship to oneself and one's relationship to others as being one of degree (cf. Ibid. 144).

Here I would like to ask the same question: Why does Korsgaard not think of arguing that the reflective structure of human discourse requires that individuals identify themselves with some law or principle that will govern their exchanging and sharing of reasons, and that this is the source of normativity? Or in short, why do others appear so untrustworthy? As G. A. Cohen points out in his commentary on Korsgaard, we should place importance on "Kant's insistence that the imperative of morals must not come from *human* nature, nor even human *reason*, should there be any respect in which human reason differs from reason as such" (Ibid. 172). Indeed, Kant strictly differentiates pure reason from our "empirically conditioned" (*CPrR*, V, 16) ones. Given that, our autonomy does not come from our *human* nature, as Korsgaard claims, but is, rather, more primordial. And then it must not be the case that *each of us has his/her own pure human reason* but, rather, that *we human beings share one pure reason*.

This is the point that I would like to emphasize in this paper. In my opinion, those first-personal interpretations of Kant adopt the former understanding and, from the fact that there seems be no guarantee of identity between every pure human reason, they are led to hold

[2] For Kant, the principle of *autonomy* is "the principle of every human will as a will giving universal laws in all its maxims" (*FMM*, IV, 432), and all other principles are principles of *heteronomy*.

that following agreements with others amounts to potentially being obligated by a reason other than my own. Of course, we know from both everyday life and human history that our reasons have a very similar or seemingly identical system but, in principle, we cannot be sure of their identity. This unwarrantedness makes no substantial difference in so far as we employ a 'minimal' interpretation of the categorical imperative, as in Wellmer.

However, I do not think that such a modest moral philosophy is satisfactory as an interpretation of Kant. On the contrary, I believe that such an understanding of pure reason and others is quite the opposite to that of Kant, considered as a transcendental idealist. I am not going to argue that such views are implausible, but in so far as those who hold them call themselves Kantian, they should be aware that they are, in fact, deviating from transcendental idealism. In what follows, I will try to justify this claim by (re)constructing Kant's unwritten theory of others in terms of transcendental idealism.

3

In the argument so far, I have tried to discover the connection between two ideas. On the one hand, the first-personal point of view is precedent over social interactions including intersubjective consensus. On the other hand, others may have different (pure) reasons than my own. These two ideas sound irrelevant to each other, but as I see it, they run along the same lines. In short, according to this view, we should not put our trust in the judgements of others, or even those arrived at with others, because they might have their own reasons. Is this at all in accordance with Kant's thought? As frequently pointed out, Kant seldom argues about others. Certainly he writes about, for example, our moral obligations to others, the realm of ends, the *sensus communis*, and the cosmopolitan law or right. But he gives, in fact, no substantial argument on the other's reasons or the reason of others.

Here I would like to cite his argument on empirical idealism and transcendental realism and apply it to the current issue. In the first edition of the *Critique of Pure Reason*, Kant introduced the contrast between transcendental and empirical idealism, and differentiated himself from the latter. Empirical idealists are those who hold that we can never, by way of any possible experience, be completely certain as to the existence of external objects (cf. *CPR*, A368-369). According to Kant, this position is the consequence of being a transcendental realist:

> It is, in fact, this transcendental realist who afterwards plays the part of empirical idealist. After wrongly supposing that objects of the senses, if they are to be external,

> must have an existence by themselves, and independently of the senses, he finds that, judged from this point of view, all our sensuous representations are inadequate to establish their reality (*CPR*, A369).

However, Kant himself is a transcendental idealist, who holds that all appearances should be understood as being only representations, not things in themselves (cf. ibid.).

I think that we can compare this anti-Kantian standpoint, namely empirical idealism based on transcendental realism, with those first-personalist Kantian philosophers, because they do not put their trust in others as partners in moral discourse or the 'exchanging and sharing' of our convictions. Let me try to model this standpoint after Kant's description. Assume that the following paraphrase is possible: that I can never be completely certain as to whether others have the same reason as me, amounts to the fact that I can never be completely certain as to the existence of other *rational* beings. This appears to be a kind of solipsism, and is obviously stronger than what the aforementioned self-proclaimed Kantians presuppose. However, this is no less than what they should maintain, if they maintain the universal validity of the categorical imperative. Because, then, my first-personal conviction or 'reflective endorsement' must be valid universally, despite there being others who can discuss the validity of my claim. So, the conclusion would be as follows: There is only one *rational* being among other beings, who can surely make a valid moral judgment, and it is me. But of course, such a presupposition would be nonsense because, given that, it would become extremely difficult to explain how such an egoistic moral theory could be adopted by other people and without collision. Thus, they had to retreat to regarding the validity of my conviction as limited, in order to coexist with other egoistic people[3].

However, why do these thinkers have to be driven to such a predicament? Because, according to our understanding of Kant, they suppose that 'the other,' in its most radical sense, must have a reason other than my pure reason. Suppose for instance Kant's concept of 'radical evil.' 'The other' in the meaning of this concept must be for our reason completely transcendent or incommensurable, so only to be imagined. However, do we really have to assume that there should be such a transcendent other outside our space of reasons? If, some day, just such an other walks into our space of reasons, then he/she would be, in fact, not external for us in the sense of 'transcendent.' Apparently, this assumption is excessive for Kantian philosophy. Strictly speaking, 'radical evil' for Kant is still not enough for such a transcendent other because it is already specified in its way of being by Kant, so it is quite

[3] Korsgaard optimistically believes that the problem of such an egoist as a 'practical solipsist' will be dismissed once we notice that our reasons for acting are private only in an "incidental or ephemeral" manner (Korsgaard [1996] 2014, 135), and "inherently sharable" (ibid.). But she could say so only in so far as she considered the source of the normativity consists, rather, in this institution of sharing itself than our personal identities.

possible to conceptualize within our space of reasons as the 'negative quantity,' which is precisely opposite to our reason.

If the presupposition of such an other is not acceptable, then we must hold that every logically possible difference among us is, as a whole, within our sole pure reason, and so within our space of reasons. In other words, we must adopt transcendental idealism concerning the other. Accordingly, this will lead us to Kantian empirical realism regarding the other, which Kant himself does not develop enough. It should describe others not by means of other reasons but by means of personal difference and private ends within the sole, pure reason. Under this condition alone, Kant's concept of the 'realm of ends' has its relevance:

> By "realm [of ends]" I understand the systematic union of different rational beings through common laws. Because laws determine ends with regard to their universal validity, if we abstract from the personal difference of rational beings and thus from all content of their private ends, we can think of a whole of all ends in systematic connection, a whole of rational beings as ends in themselves as well as of the particular ends which each may set for himself (*FMM*, IV, 433).

And finally, our understanding of pure reason should be reformed in terms of intersubjective concepts, as it has allowed those Kantians to regard it as *human* and, then, to be easily led to the familiar first-personal anti-Kantian moral philosophy. We should make it clear that our pure reason is not *human* reason, which could be inferred from the first-personal point of view, but an interpersonal and intersubjective one. For example, Korsgaard's 'exchanging and sharing' may work in this way. However, there is a more systematically elaborate concept available for us, called 'discourse', which embodies communicative or discursive reason.

Abbreviations for Kant's Writings

CPR = *Critique of Pure Reason*, trans. by Norman Kemp Smith, intr. by Howard Caygill, bibliography compiled by Gary Banham, Palgrave Macmillan, 2007.

CPrR = *Critique of Practical Reason*, trans. and ed. by Mary Gregor, intr. by Andrews Reath, 11th pr., Cambridge University Press, [1997] 2010.

FMM = Foundations of the Metaphysics of Morals, in: *Critique of Practical Reason And Other Writings in Moral Philosophy*, trans. and ed. with an intr. by Lewis White Beck, The university of Chicago Press, 1949, pp. 50-117.

References

Habermas, Jürgen (1990), *Moral Consciousness and Communicative Action*, trans. by Christian Lenhardt and Shierry Weber Nicholsen, intr. by Thomas McCarthy, The MIT Press.

Korsgaard, Christine M. ([1996] 2014), *The Source of Normativity*, 20. pr., Cambridge University Press.

Wellmer, Albrecht (1991), *The Persistence of Modernity*, trans. by David Midgley, Polity Press.

Yoshime, Michihito (2016), "Consequences of the Transcendental-Pragmatic Consensus Theory of Truth", in: Halla Kim/Steven Hoeltzel (eds.), *Transcendental Inquiry. Its History, Methods and Critiques*, Palgrave Macmillan, pp. 263-283.