Published in: European Journal of Philosophy 18, 2, June 2010, pp. 296-310.

Review Article

The Moral Standpoint: First-Personal or Second-Personal? [1]

Herlinde Pauer-Studer
University of Vienna

The Second-Person Standpoint: Morality, Respect, and Accountability, by Stephen Darwall. Cambridge, MA. and London: Harvard University Press, 2006, xii + 348 pp.
ISBN-13: 9780674022744 hb £41.95; ISBN-10: 0674022742 pb £16.95

Stephen Darwall's *The Second-Person Standpoint: Morality, Respect, and Accountability* is a highly stimulating and impressive book. Its main goal is to give an account of morality in terms of the second-person standpoint. Morality is, as Darwall defines it, a matter of making and acknowledging 'claims on one another's conduct and will' (p. 3). The validity of demands addressed by one person to another depends on normative relations between them, i.e., whether the one has the legitimate authority to hold the other accountable.

The second-person standpoint gives rise to second-personal reasons. Second-personal reasons are relational and agent-relative. A form of reciprocal respect is part and parcel of all second-personal reason-giving. Dignity is defined as the *de jure* authority[2] persons have 'to demand certain treatment of each other' (p. 13). The second-person account presupposes that you and I share a common normative ground: we have second-personal authority, second-personal

[1] For critical comments on an earlier version of the paper I would like to thank J. David Velleman. An earlier version of the paper was presented to the First Annual Dutch Conference in Practical Philosophy, Oct. 2009. I would like to thank the participants, especially Stephen Darwall, for helpful discussion.
[2] The authority to make a legitimate claim or demand is not de facto, but legitimate authority.

competence, and accountability as free and rational agents. Demands addressed under these constraints meet the relevant 'normative felicity conditions' (pp. 74-76).

For Darwall, second personal address is connected with reactive attitudes like resentment, blame, indignation and guilt. He considers the reactive attitudes, as they are outlined by Strawson, to be indicators of what can be rightfully demanded of others. They are the correct response if others do not recognize the legitimacy of certain claims.

Darwall illustrates the making of demands on the basis of second-personal reasons with two recurring examples. Suppose someone steps on your foot, causing you pain. Then you are according to Darwall in a normative position to address to the other person a claim to remove her foot. You give the other person a second-personal reason to do so (pp. 6-8). If the other person does not comply with your demand, then your reaction of blame or feeling resentment would be justified.

The other example is the case of a sergeant giving orders to his platoon. The sergeant, in his professional role and authority, gives his subordinates a second-personal reason to comply by addressing an order to them. The reason is created by the normative standing of the sergeant vis-à-vis his platoon. The sergeant 'addresses a reason that would not exist but for her authority to address it through her command. Similarly, when you demand that someone move his foot from on top of yours, you presuppose an irreducibly second-personal standing to address this second-personal reason' (p. 13).

Darwall sees at the core of his account four interrelated ideas: claim, accountability, second-personal reason, and second-personal authority. These notions form, as he explains, 'an interdefinable circle, each implies all the rest' (p. 12). Moreover, the foundations for these concepts cannot be gained from evaluative or normative concepts which are not second-personal.

Darwall rejects a third-person standpoint as developed by Thomas Nagel, namely, regarding others from an impersonal agent-neutral perspective (p. 102). The main reason is that for Darwall a third-person standpoint creates merely an epistemic relation to reasons, not a specifically normative one addressing and engaging another person's will by a legitimate claim. Take the example of stepping on another person's foot. One might, Darwall concedes,

claim that the case gives rise to two different kinds of reasons: an agent-neutral one resulting from the badness of pain, and an agent-relative one which is based on your authority to demand that the other person not step on your foot. An agent-neutral point of view, however, constitutes an objective, detached perspective on the state of the world. By appealing to agent-neutral reasons you assume according to Darwall epistemic authority to direct the other person's beliefs about practical reasons, but you do not address her will by pointing out a normative commitment she has towards you.[3] The specifically normative element would be missing.

Darwall equally rules out a first-person account of morality that locates the source of normativity in the singular self-legislation or singular reflective endorsement of a rational subject, independent of any second-personal reference. Again, his point is that normativity is a matter of second-personal reason-giving. As he tries to show by drawing on Fichte's practical philosophy, the idea of an 'agent's own self-determining choice' is dependent on the normative authority of others to address a claim or summons (*Aufforderung*) to the agent and therefore presupposes a 'mutual second-personality that addresser and addressee share' (p. 21). Darwall does not reject a first-person standpoint per se. He rejects only those versions of a first-person standpoint which do not include a second-personal aspect. In fact, he claims the second-person standpoint to be a version of the first-person standpoint, since second-personal reasons are agent-relative.

The general picture of morality that Darwall offers seems very attractive. Moral philosophers, especially feminist philosophers, have forcefully criticized a 'view from nowhere'-perspective as inappropriate for morality. Darwall's theory apparently meets the requirement that the main focus of morality should be our relations to 'concrete others' and not a detached God-like perspective on the world. His approach also avoids the apparent solipsism that haunts an account of morality in terms of the isolated reflections of *the* rational subject. Since Darwall attributes a specific role to the reactive attitudes of blame, indignation, and resentment, the voice of others receives a considerable place in our moral deliberations.

[3] Darwall objects that Nagel's grounding of all normative reasons in an agent-neutral and impersonal perspective makes it impossible for Nagel to give an appropriate account of agent-relative reasons (see Darwall 2007: 53, 54). Darwall does not want to eliminate the category of agent-neutral reasons and considerations. His account of the relations between agent-neutral and agent-relative reasons is that in addition to an agent-neutral reason resulting e.g. from the badness of a state, an additional second-personal reason is created by the idea of a warranted claim the person has towards another person. Nagel, he argues, misses this specific form of a second-personal reason.

There has already been illuminating discussion of Darwall's book.[4] For the most part the debate has centred on analyzing the notion of second-personal reasons and assessing whether Darwall's claim concerning their centrality is tenable. An issue that has received less attention, and on which I am going to focus, concerns the structure and details of Darwall's foundational programme. Darwall accepts the Categorical Imperative (CI) as the fundamental principle of morality, gives it a second-personal interpretation and in this way provides a contractualist version of moral theory.

In the first part of my discussion I address the question whether Darwall's reformulation of Kantian ethics within a second-person standpoint theory seems more plausible than the interpretation of Kantian moral theory in terms of a first-person internalism. In the second part I will take a closer look at Darwall's account of practical reason and autonomy. I close with some comments on the commitments of Darwall's contractualism.

Darwall's Contractualism

Should we prefer Darwall's second-personal reconstruction of the Kantian programme to a first-person interpretation? What exactly is the difference between the two readings? What does an interpretation in terms of the second-person standpoint tell us about the meaning of the various CI-formulas and their structural relatedness?

The most elaborate reading of Kantian ethics in terms of a first-person standpoint has been developed by Christine Korsgaard (Korsgaard 1996a, 1997, 2009).[5] She reconstructs Kant's theory as a form of 'constitutive internalism'. The principles of practical reason - the instrumental principle and the Categorical Imperatives - are constitutive of the person as a rational agent. The self-legislation of an autonomous will is the source of normativity. The capacity to be an autonomous agent - an agent recognizing the force of universality and

[4] See the 'Symposium on Stephen Darwall's *The Second-Person Standpoint*', *Ethics*, 118, 1, 2007: 8-69, containing the following articles: Korsgaard (2007), Wallace (2007), Watson (2007) Darwall (2007).
See also 'Kritik und Antwort zu Stephen Darwall: The Second-Person Standpoint', *Deutsche Zeitschrift für Philosophie* 57 (2009): 159-179 including the following articles: Betzler 2009, Rödl 2009, Schaber 2009, Darwall 2009.

[5] Not all philosophers have defended a first-person reading of Kant's moral theory. Schopenhauer, for example, has interpreted Kant's theory as a third-person standpoint account in which pure reason just replaces God's impersonal perspective. Ernst Tugendhat follows Schopenhauer's reading and criticizes Kant's ethics as relying on an impersonal *Vernunft fettgedruckt* which has no link to agent-relative reasons. Cf. Tugendhat 1993. I leave that reading of Kant aside.

valuing her or his humanity as an end in itself - is indispensable for having reasons for action at all.

Darwall's claim is that all Kantian strategies for grounding the Categorical Imperative (the Formula of Humanity, the Formula of Universal Law, the idea of the Kingdom of Ends) — including the strategies of Kant himself and those of Neo-Kantians like Korsgaard— fail because the Imperative's normative force cannot be established within a first-person deliberative perspective. This is a bold argument and it needs some attention to Kant's attempts to justify the Categorical Imperative in order to assess it.

To recall: In the *Groundwork of the Metaphysics of Morals* (*GMM*) Kant tries to provide a deduction of the Categorical Imperative (CI); the argument runs from freedom of the will and autonomy to the moral law. In the *Critique of Practical Reason* (*CPR*) Kant starts from the premise that we are conscious of the moral law, the Fact of Reason, and that our consciousness of the moral law makes us aware that we are free.

Kant illustrates the Fact of Reason with the example of someone who is demanded by his prince, on pain of execution, to make a false statement against an honourable man. Kant argues that the agent must admit that, his love of life notwithstanding, it would be possible for him to refuse the prince's demand, which shows that the agent recognizes that he can do an action because he is aware that he ought to do so. Thereby the person recognizes that he is free – something which, Kant adds, 'without the moral law, would have remained unknown to him' (Kant 1997: 28; Academy edition 5:30).

There has been an intricate and ongoing debate among Kant scholars about what exactly Kant's Fact of Reason-argument establishes.[6] Different interpretations have been offered as to whether Kant's reformulation of the argument in the *CPR* means that he gave up on the project of a deduction of the CI as undertaken in the *GMM* or whether he merely tried to improve his argument by establishing that we are free, thereby removing his own earlier doubts that the deduction in the *GMM* was circular because it simply assumed that we are free.

[6] For an excellent view of the discussion see Kleingeld (forthcoming 2010).

Darwall follows the line that Kant dismissed the deduction in the *GMM* and offered a distinct argument in the *CPR*. He argues that Kant was correct to abandon the deduction in the *GMM* because it is indeed circular: Kant's argument just presupposes that we are autonomous. Thus, the only successful attempt can be to follow the Fact of Reason-argument. According to Darwall, however, the adoption of the second-person standpoint is required to vindicate Kant's Fact of Reason-argument.

Darwall objects that Kant's argument in the *CPR* merely suggests the moral law as an 'open deliberative alternative' (p. 240). That the agent is aware that he can refuse the prince's demand is too weak. Something stronger is needed, namely, that the person *ought* to refuse the prince's demand. Darwall claims that the missing link is provided by the second-person standpoint. If the agent *ought* to refuse the demand of the prince, then she is accountable and responsible for doing so. Henceforth the moral community has the authority to demand that the agent refuse. The normative felicity conditions are met – and blame can be put on the person. As Darwall writes:

> Thus, although there is no conceptual pressure to hold that the mere existence of good and sufficient reasons for someone to do something entails that he knows or even can know this (would that it were so!), it does seem to be a conceptual requirement of blaming and holding Citizen responsible for not refusing the prince's demand (if she fails to) that we presuppose that she must have been in a position to know that she should have refused and that she could have determined herself to refuse by the relevant reasons. (p. 241)

But what, we might ask, is the contribution of the second-person standpoint? Does it make a specific difference to the reasoning of the agent? Take again Kant's example in the passage on the Fact of Reason. For Darwall it requires a form of CI reasoning - including a specific second-personal element - for the agent to become aware that he ought to refuse the prince's demand. This second-personal element is cashed out as the second-personal competence we attribute to the agent.

Maybe I do not grasp the full depth of Darwall's argument, but what he presents seems not enough to establish a specific second-personal form of reasoning. His point is well taken that we as members of the moral community have to attribute to an agent a competence to reason and deliberate correctly in order to hold her accountable and responsible. But making explicit the preconditions of blame and responsibility does not show that the reasoning process as such must be second-personal.

The legitimacy of the reactions of the moral community – the legitimacy of holding the agent accountable and responsible in case he fails to refuse the prince's order - depends on the fact that the agent *ought* to have done so. But that he ought to do so is, I think, simply presupposed by Darwall, not shown to be the outcome of second-personal deliberation, as the following passage shows:

> It is common ground that Citizen is morally obligated to refuse the prince's demand, that it would be wrong for her not to do so, and that that is therefore what she should do and, consequently, what she can do. (p. 240)

Kant presupposes in the passage on the prince's demand that the agent recognizes that he ought to refuse – but it is individual CI reasoning that leads the agent to that conclusion. On which specifically second-personal reasoning would agents have to rely in order to find out what is correct? The claim that we reason together by addressing demands to each other cannot do the job on its own. The discourse would, I think, have to be structured in light of specific second-personal principles. Darwall, however, does not point out any principles, other than the classical Categorical Imperative-test, which would characterize a second-personal deliberation. In a reply to his critics he even affirms that 'being governed by a formal principle like the Categorical Imperative (CI) is the form that moral reasoning would have to take if it is to lead us to conclusions that we can be held responsible for reaching' (Darwall 2007: 58).

This raises the question whether there is any substantial difference between a first-person and a second-person interpretation of Kant's moral theory. One might argue that, given the structure it has, the Categorical Imperative test requires me to consider the claims of others; its application therefore trivially presupposes something like 'second personal competence'. The Categorical Imperative procedure requires me to ask what I can do under the condition that others could act in the same way and could consent to my maxims of treating them. Seen this way the distinction between first-personal and second-personal reasoning more or less vanishes. The only way to make it out at all would be to point to the difference between an 'internal dialogue' with others and a direct argument with others in real life about their claims. But this can hardly be the site of a relevant distinction: moral thinking simply cannot be reduced to real verbal encounters with others about their legitimate claims.

The difference between a first-personal and a second-personal account of morality becomes clearer when we turn to Darwall's reformulation of Kantian moral theory as a form of contractualism. Darwall's foundational programme is: Take the Formula of Humanity (FH) as fundamental, interpret FH in terms of the Formula of the Realm of Ends (FRE), and then interpret the Formula of Universal Law (FUL) in its light (pp. 304-309, esp. p. 308). In more detail: The starting point is the dignity of persons, the idea of treating each other as ends and never merely as means (FH). The concept of dignity has to be spelled out in second-personal terms, namely, in terms of mutual accountability of equals. This brings us to the idea of a realm or kingdom of ends (FRE), a community of rational beings united by common laws requiring us to treat one another as ends and never merely as means. Dignity commits us to addressing others with second-personal demands that cannot be reasonably rejected and to which free and rational agents hold themselves accountable. FUL just specifies what this means in regard to the particular will and reasoning of the individual person. The idea of the equal recognition of others excludes regarding the individual person as having a special standing – an idea which is spelled out in asking whether one's maxims could be thought or willed as universal laws.

There is in fact a congruence between Darwall's programme and Korsgaard's reconstruction of the connections between the different formulas of the Categorical Imperative in *The Sources of Normativity*. Darwall's account mirrors exactly what Korsgaard wants to achieve. She first wants to make the step to the moral law (the demand to live together in the Kingdom of Ends, 'the republic of all rational beings');[7] then to interpret the FH as a specific interpretation of the moral law in regard to human beings; and then to establish the FUL as the formal counterpart of the idea of treating others as ends-in-themselves.

Korsgaard is often remarkably close to a contractualist account. For example, in the passage in which she introduces the idea of the 'moral law' in *The Sources of Normativity* she talks in the language of contractualism:

> The moral law, in the Kantian system, is the law of what Kant calls the Kingdom of Ends, the republic of all rational beings. The moral law tells us to act only on maxims that all rational beings could agree to act on together in a workable cooperative system. (Korsgaard 1996a: 99)

[7] Korsgaard draws, otherwise than Kant, a distinction between the moral law and the CI. See Korsgaard 1996a: 99.

Why does Korsgaard not simply subscribe to a Kantian version of contractualism as a foundation for morality? I suppose two reasons are pertinent, both of which are versions of the worry that contractualism does not answer to the specific requirements of a theory of ethics.

The first reason is that Korsgaard sticks to Kant's foundational programme in the *GMM*, which works from autonomy of the will to the Formula of Universal Law (FUL) and from there to the Formula of Humanity (FH). Kant, she argues, did not give up his justification (deduction) of the Categorical Imperative in the *Groundwork* because he suspected it to be circular. According to her interpretation Kant was merely worried that he had tacitly presupposed in the *GMM* that *we can be motivated* by the CI without providing an argument for that assumption. Therefore, Kant tried to remedy that problem by introducing the Fact of Reason-argument, which shows that we can be motivated by the moral law (Korsgaard 1996c).[8] Korsgaard considers Kant's step from autonomy to the CI in the *GMM* as valid. Her own justificatory argument is (following closely Kant's argument in the *GMM*): A free will or an autonomous will acts according to its own principle or norm, that is to say, it is guided by a self-given law. The principle of a free will is henceforth a law, and this condition, of being a law, is exactly fulfilled by the Categorical Imperative. Korsgaard considers the acceptance of that argument as indispensable for a foundation of ethics on Kantian grounds.

Here there arises a problem, however. Since Korsgaard explicitly defends the deduction of the CI in the *GMM*, she seems committed to making first the step to FUL and from there to derive FH. The move, however, from a formal principle like the FUL to a substantial principle, the FH, seems notoriously difficult and hardly defensible.

As mentioned, Korsgaard explicitly favors the other strategy, namely, to get first to the moral law, and then make the step to the FH and then to the FUL (Korsgaard 2002). So she provides another argument, the argument from the value of humanity: To be an agent we need a normative structure, a practical identity. But we cannot develop practical identities if we do not attribute value to ourselves, if we do not value our humanity. And to value our humanity

[8] I would like to leave aside the question whether Darwall's or Korsgaard's reading of Kant is correct. I consider Korsgaard's motivation-based interpretation of the Fact of Reason-argument as a highly interesting reading, but it is, in my view, part of her own original account of a Kantian theory and less what Kant had in mind.

we must equally value the humanity of others. The publicity of reason forces us to consider the humanity of others as valuable if we consider ourselves as valuable.[9]

Even if we accept Korsgaard's argument that to be an agent entails that you must value yourself as a person and, because of the publicity of reasons, also value other persons, that argument by itself does not imply that you must value the persons around you in the specific and demanding way that is prescribed by Kant's idea of treating others as ends in themselves. I can take myself as important and value myself also by following the principle that my interests should be simply prior to those of others – and the publicity condition leads us in this case to make concessions, but not to the deep form of respect for others that Kant had in mind. Moreover, it seems difficult to imagine why I should be denied identity as an agent because of making an egoistic strategy my principle.

At this point a contractualist or relational model, as Darwall proposes it, seems indeed more plausible. We are brought directly to the moral law by looking for an answer to the question 'On what principles must our relations to each other be based to which we all as free and rational agents can reasonably consent?'. The contractual aspect is to work out the normative presuppositions of a community united by 'common laws' that no one can reasonably reject.

How well does a contractualist model of morality fulfil the specific requirements of individual morality? This leads us to the second reason why Korsgaard might have some reservations about a contractualist framework.

Contractualism has a tendency to blur the distinction between individual morality and public morality. To ask which principles we cannot reasonably reject does not make clear whether we are referring to ethical principles or principles of justice. Equally the question which claims of others we cannot reasonably reject does not specify whether we should assess those demands on ethical grounds or grounds of justice. There is, however, an important distinction here – at least if we follow Kant and Rawls.

[9] A version of this argument works via the value of ends: In order to be agents and have a reason to act we must consider our ends as important and put value on our ends. But we can value our ends merely by valuing ourselves. But if I have reason to value myself then I have reason to value all others. Darwall criticizes the value of ends-argument and the value of agents-argument: they would only work if the step from autonomy to the moral law would work, but this step cannot be guaranteed within a Kantian approach since the argument in *GMM*, part 3, is circular and fails (pp. 229-235). This leaves open the question whether Darwall considers the arguments as valid within the second-person standpoint account since the step from autonomy to the moral law is guaranteed on second-personal terms.

Kant's political philosophy is based on contractualism. Kant draws a clear line, however, between the sphere of external freedom (justice, law) and the sphere of internal freedom, the sphere of convictions, volitions, intentions, and motives. The latter is the subject of ethics, and the relevant moral constraints in regard to one's convictions, volitions, and motives must be set by one's own conscience, i.e., one's first-person assent and motivational internalization of the moral standards. In the realm of external freedom, the ethical aspect of your particular reasons and motivations plays no role. For Kant it is crucial *that* people follow the principle of right which obligates them to respect the equal external freedom of others; the motivational reasons *why* persons comply with the principle of right are according to Kant irrelevant (Kant 1996: 23-34). In Rawls's theory the motivational neutrality of the standards of justice is guaranteed by applying the principles of justice to institutions and not to individual behaviour.

I think that Korsgaard's acceptance of the sharp distinction in Kant's practical philosophy between ethics and public morality (justice, law), between inner and external freedom, prevent her from situating all of morality in a contractualist framework. A central thesis of Korsgaard's is that a proper ethical theory must provide a close connection between normativity and motivation. That Kant's theory relies on 'a *motivational analysis* of the notion of duty or rightness' (Kant 1996d: 60) to elucidate the principle of a good will is a main reason why Korsgaard follows the line taken in the GMS. Obligation and motivation, she writes, do not fall apart in the *GMM*, because a good-willed person 'does the right thing because it is the right thing' (Kant 1996d: 60). The programme of ethics, namely, to find the principle of a good will, commits us to first-personal deliberation and assent.

That Darwall does not give enough attention to the important distinction between individual morality and public morality becomes apparent in the way he relies on aspects of Fichte's practical philosophy to elucidate and support his second-person standpoint account. Darwall argues that Fichte, in contrast to Kant, explicitly brings the second-person standpoint and second-personal address to bear on moral theory.[10] Darwall draws, interestingly enough, on Fichte's philosophy of right, and not on Fichte's ethics, the *Sittenlehre*, to make his point. Darwall specifies Fichte's form of second-personal address as making and acknowledging a summons in terms of Fichte's principle of right which reads: 'I must in all cases recognize the

[10] What Darwall calls 'Fichte's Point' is that two conditions must be satisfied for second-personal address and for making a summons (*Aufforderung)*: a shared equal authority and a shared freedom 'to act on claims that are rooted in this authority' (p. 246).

free being outside me as a free being, i.e. I must limit my freedom through the concept of the possibility of his freedom' (Fichte 2000: 49). But this entails merely that I must recognize that the limits of my external freedom are set by the existence of others and their claim to equal external freedom.[11] In *The System of Ethics* Fichte clearly privileges an internal and first-personal point of view. The self-determination of the will of the thinking subject is the basis of the deduction of the moral law (Fichte 2005: XXI, 19-63; 98-116).[12]

There is an additional aspect which shows that Darwall cannot and in fact does not dispel a first-person standpoint. An essential element in Darwall's contractualist theory is responsibility and accountability to others. However, that second-personal aspect must have a first-person counterpart. What Darwall calls *Pufendorf's point* is relevant here: If we, as members of the moral community, hold another person responsible for complying with a moral obligation, we take it that the person can hold herself responsible. In Darwall's own words:

> To intelligibly hold someone responsible, we must assume that she can hold herself responsible in her own reasoning and thought. And to do that, she must be able to take up a second-person standpoint on herself and make and acknowledge demands of herself from that point of view. (p. 23)

---

[11] The main purpose of Darwall's appeal to Fichte is to use Fichte's remarks about external freedom to show that inner freedom must include second-personal competence. As Darwall writes: 'Fichte believes that second-personal engagement commits addresser and addresse alike to limiting their "external freedom" through the 'principle of right'. In so doing each "lets his own external freedom be limited through inner freedom"' (p. 254). The relevant passage in Fichte reads: 'It turns out that, in thought, each member of the community lets his own external freedom be limited through inner freedom, so that all others beside him can also be externally free' (Fichte 2000: 10). But this establishes merely the trivial point, already discussed, that of course the deliberations of rational agents must take into account the existence and legitimate claims of others. As already mentioned, second-personal competence does not show that a second-personal reasoning process is inherently different from a first-person deliberation taking into account the claims of others as rational and free agents.

[12] 'Ethics' is defined by Fichte as 'the *theory of our consciousness* of our moral nature in general and of our specific duties in particular' (Fichte 2000: 21) (italics in the original). Fichte introduces an interpersonality thesis in the *Wissenschaftslehre*. The experience of others and their summons is the basis of respecting the separateness and self-reliance (*Eigenständigkeit*) of others, which presupposes to acknowledge the self-determination of one's own will. So again nothing beyond the first-person standpoint dealing with the presence of others and their summons is presupposed. (Cf. Klotz 2002: 159-169) Klotz writes: ‚Jede Person, so läßt sich der Grundgedanke in Fichtes Individualitätskonzept wiedergeben, ist vor die Aufgabe gestellt, die ihr wesentliche Ausrichtung ihres Wollens inhaltlich selbst zu bestimmen….Von einem für freies Handeln grundlegenden Selbstverhältnis ist demnach im Sinne eines reflektierten Bewusstseins der voluntativen Ausrichtung zu sprechen, die für die je eigene Identität konstitutiv ist. Kraft dieses Selbstverhältnisses verfügt eine Person über eine Orientierung, die ihre bewussten Entscheidungen durchgängig leitet und zugleich „beschränkt"' (Klotz 2002: 167, 168).
In English (my translation): 'To set out the basic idea of Fichte's concept of individuality: Each person is confronted by the task of determining for herself the content and direction of her volitions….Therefore to consider the relation to one's self as crucial for acting freely has to be interpreted in terms of a reflective consciousness of volitional direction which is constitutive of one's own identity. By virtue of that relation to one's self a person possesses an orientation which guides, but equally „constrains", her deliberate choices.'

What is crucial is that the person must rely on her *own reasoning and judgement*, and not simply be driven by her fear of sanctions from others. Just as Pufendorf claimed that moral obligations derive not merely from the external authority of God threatening us with sanctions (in case we violate moral obligations), but from our understanding of God's demands, so our commitment to moral obligations is due to our understanding of the demands which we, as rational agents, can address to ourselves. Darwall points out that there is a difference between 'coercion, on the one hand, and free self-determination by an internal acceptance of an authoritative demand, on the other' (p. 23). So Darwall presupposes internalism on the part of the individual subject: the agent acknowledges the force of obligations. Hence, even if the foundation of the moral law rests on a second-personal contractualist agreement with others, there must be a corresponding first-person source of normativity.

I argued that a contractualist account, as Darwall proposes it, offers a direct route to the moral law. The moral law expresses the general normative idea of a community of equals – constituted by principles which cannot be reasonably rejected. However, Darwall needs to further restrict that general idea in order to make room for the crucial distinction between a theory of justice and ethics. How could that distinction be made? The plausible way would be to distinguish more clearly the problems to which the principle of right is the answer on the one hand, and on the other hand the more particular ethical versions of the Categorical Imperative, namely the Formula of Humanity and the Formula of Universal Law are the answers.[13] The idea of principles and demands which cannot be rejected can be spelled out in more nuanced ways. The principle of right is the guideline specifying which claims cannot be rejected on grounds of justice respecting the equal external freedom of others. The Formula of Humanity and the Formula of Universal Law would be the guidelines to specify which demands of others cannot be reasonably rejected, constraining thereby the maxims on which we are morally allowed to act. But all this does not relativize the point made earlier that on the

[13] The question how to interpret the architectonic of Kant's practical philosophy has often puzzled interpreters. Kant does not formulate a general Categorical Imperative from which the CI (individual morality) on the one hand and the principle of right (public morality, justice) on the other hand is deducible. It is also not clear how such a general CI would have to be formulated in order to capture the distinction between ethics and the philosophy of right: emphasis on inner attitude and motivation on the one hand, motivational neutrality on the other hand. It seems more appropriate to understand Kant's practical philosophy as focused on presenting two distinctive regressive arguments (one in respect to ethics, the other in respect to the sphere of right) which provide the ethical formulas of the CI (i.e. FUL, FH) and the principle of right as the answers to the specific central questions of ethics and the philosophy of right. These questions are: In regard to ethics: What is the principle of good action? In regard to the sphere of right: What justifies coercion?
One might object that Kant does not leave room for a search for principles of justice as standards of public morality, functioning as guidelines for the sphere of law and the legal design of the basic institutions of society. But such principles of justice could equally be reconstructed as the answers to the question: Which form of society would free and rational agents choose who want to pursue their conception of the good life?

level of individual ethics a sharp distinction between a first-person and a second-person standpoint can hardly be drawn. The assessment of one's maxims includes the recognition of the demands of others which cannot be reasonably rejected.

To conclude: A relational or contractualist theory does seem to be a more straightforward way to get to the moral law and the structure of the different formulas. However, contractualism does not do away with the internal perspective and first-person deliberation.

Autonomy, Morality, and Transcendental Commitments

There is a deep ambivalence in Darwall's position. On the one hand, Darwall claims not only to save the strength of the Kantian foundational programme but to establish it in the only proper way by providing a second-person standpoint interpretation. On the other hand, he backs away from the strong commitments of Kant's moral theory. This ambivalence becomes obvious in Darwall's treatment of practical reason and autonomy as well as in his discussion of the 'transcendental presuppositions' of the second-person standpoint.

Let us first look at practical reason and autonomy. Some passages read as if Darwall thinks Kant's strong conception of autonomy – that the relation between autonomy and the Categorical Imperative is analytic - to be implausible. Take Darwall's remark: '[N]othing in the bare project of acting for reasons, first-personally, commits a deliberating agent to autonomy as Kant defines it' (p. 214). Darwall explicitly rejects Kant's claim that 'in presupposing autonomy a rational agent is committed to the CI (the moral law)' (p. 30). Darwall, moreover, concedes that 'autonomy in the familiar sense' just means the ability to take a critical point of view on one's desires and beliefs. The capacity to step back and to revise them, if critical reflection demands, constitutes us as agents (p. 228). Interestingly enough, Darwall dismisses this picture of agency and deliberation as 'naïve' and inappropriate.

Darwall takes the example of a 'naïve practical reasoner' to illustrate the weakness of Kant's conception of practical reason. He objects that the naïve conception of deliberation cannot be ruled out from the first-person standpoint of deliberation. It needs a second-person standpoint to 'dispel the naïveté' (p. 217).

In more detail: In the case of the naïve reasoner there is according to Darwall a structural analogy between theoretical and practical reasoning. In theoretical reasoning the premises are putative facts about the world as they seem to the reasoner from the perspective of his beliefs; in the case of practical reasoning the starting point are premises about how it would be desirable for the world to be, seen from the perspective of the reasoner's desires. The naïve model of reasoning therefore does not depict clearly in which way practical reasoning is special.

The same problem, Darwall argues, affects Kant's conception of practical reason: Kant does not specify any difference between theoretical and practical reasoning. Kant defines freedom as 'being free from alien causes' and being free to adopt a principle or norm. Darwall objects that this conception of negative and positive freedom 'is true of reason in all its employments. Whenever we make normative judgments concerning what there is reason to believe, feel, or do, we must presuppose negative and positive freedom in the sense that our judgments are free of alien causes and in accordance with rational norms' (p. 224). Thus, Darwall concludes, Kant has not shown that 'autonomy' is something special to practical reason.

In a way this is a strange and puzzling argument. The special element of practical reason Kant depicts is that autonomy of the will includes the moral law, the Categorical Imperative. So what exactly is it that Darwall considers deficient in Kant's position?

The crucial phrase in Darwall's claim that nothing commits a deliberating agent who is acting for reasons to Kantian autonomy, is '*first-personally*'. Darwall's main objection is that Kant, *because he adopts a first-person standpoint*, cannot show that autonomy entails the moral law. In other words: Autonomy for Kant means negative freedom (freedom from alien causes), but also positive freedom. A will is positively free if it has the capacity to adopt its own principle. However, Darwall argues, the positive principle a free will adopts can be something other than the CI; a free will can, for example also adopt an act-consequentialist principle which states that it is best for all rational agents to promote good or desirable outcomes. Kant, according to Darwall, cannot rule out the consequentialist alternative. The adoption of the CI remains merely an option on Kant's account (p. 226).

Darwall does not draw into question the connection between autonomy and morality as Kant defines it; his point is that the relation between autonomy and the Categorical Imperative

needs to be reconstructed in second-personal terms. We have to turn to second-personal address to locate the distinction between theoretical and practical reasoning. Only second-personal practical authority, Darwall explains, 'has no clear analogue in theoretical reasoning about what to believe. There can be no fundamentally second-personal reasons for belief, and hence there is no theoretical standing that is fundamentally second-personal' (p. 287). Thus the close connection between positive freedom and the Categorical Imperative can only be established within the second-person standpoint. Only if we are addressed by the 'summons' of others, and our second-person competence is engaged, does anything like 'autonomy of the will' come into the picture, because as second-personally competent agents 'we presuppose the autonomy of our respective wills' (p. 290).

In fact, Darwall tries to strengthen Kant's account. Darwall's crucial thesis is that the second-person standpoint establishes the connection between autonomy and the Categorical Imperative in a better way than Kant. Only within a second-personal account Kant can avoid the mistake of establishing the moral law merely as 'an open deliberative alternative' (p. 240). If, however, Darwall argues (as he does) that from a second-person standpoint autonomy necessarily entails the moral law, then he must also endorse the thesis that on the side of the individual agent autonomy entails the moral law. A necessary presupposition of Darwall's contractualism is that the agent must accept the second-personal demands of others from an internal normative perspective. But that is only possible if, from *a first-person standpoint*, autonomy entails the moral law.

This means that Darwall is committed to accepting Kant's strong notion of autonomy, namely, that the meaning of autonomy is cashed out as the Categorical Imperative and that the defining feature of practical reason is the Categorical Imperative. By subscribing to such a strong and moralizing conception of autonomy and practical reason (even if from a second-person standpoint) Darwall's account invites the objection, often raised against Kant, that as a result it becomes hard to understand how bad action is volitionally possible. If autonomy entails morality, then how can someone autonomously choose to do the bad?[14]

[14] Korsgaard is aware that the problem of bad action poses a challenge for her constitutive internalism. If the Categorical Imperative is constitutive of agency, then how can an agent wilfully choose the bad? Korsgaard sees resources in the constitutional model for dealing with the problem of bad action: the will can adopt the wrong law. The will can be structured by a principle which is formally a law, but which is substantively wrong or unjust. The example by which she tries to show how the soul can be governed by the wrong law is taken from Jane Austen's novel *Emma:* Harriet chooses voluntarily to be governed by Emma's will. So there is an autonomous choice for something wrong (see Korsgaard 2008: 162, 163). Korsgaard,

The problem comes up again with respect to the relational presuppositions of the second-person standpoint. Does Darwall, we might ask, necessarily presuppose a moralized conception of social relations – we necessarily address each other as equals with a shared dignity - which makes deviations from that standard impossible or pragmatically incoherent?

Darwall claims, for example, that from a second-person standpoint, dignity of persons and autonomy of the will entail each other. And he adds that '[b]oth are transcendental conditions for the very possibility of second-personal reasons (or 'normative felicity conditions')' (p. 277). Yet Darwall clearly does not want to go as far as claiming that all forms of second-personal reason-giving necessarily presuppose shared dignity.

That Darwall backs away from the commitments of transcendentalism becomes obvious when considering his account of evil practices. Darwall is sensitive to the issue that his argument might be too strong. He writes: 'When we think about familiar cases of subjection and domination that take an apparently second-personal form, it can seem quite incredible that second-personal address must presuppose anything remotely like a shared dignity' (p. 265). Darwall states that it would indeed be a failure if his line of reasoning had the consequence that bad actions or involvement in bad practices such as slavery would commit us to 'some sort of pragmatic contradiction' (p. 265).

To avoid that conclusion Darwall makes a substantial concession. He assumes that slaveholders need not accept that their slaves have equal standing: they can talk abusively. In addressing other persons besides his slaves and defending his practice to them, a slaveholder is not committed to acknowledge the dignity of his slaves. Only if the slaveholder directly addresses his slaves and tries to justify in a direct personal encounter with them that he considers them to be his property, 'then this second-personal address would commit him to the presupposition that he and they share an equal (second-personal) standing just as free and rational persons' (p. 267). Darwall explains: 'My claim, again, is only that any address of a second-personal reason, including any from a master to a slave, is committed to the presupposition that addresser and addressee share an equal normative standing as free and rational persons' (p. 268). That way of arguing obviously reduces moral reasons to direct

---

however, switches here to a more moderate concept of autonomy, namely one which does not entail the moral law.

second-personal reason-giving. Moral reasons, however, do have weight independently of a direct address or claim to others.[15]

Actually Darwall gives a quite convincing description of evil action when he describes the case of Stalin. Stalin was eager to justify his brutal murders by finding deficiencies in his adversaries so that his condemnation and cruel reaction seemed intelligible. Darwall writes that 'a justified authority over others seemed manifest to him, justified in ways that, as it seemed to him, others should be able to appreciate. And even his cruelest murders were accompanied, indeed fueled, by self-justifying emotions and narratives. It seems no exaggeration to say, in fact, that Stalin's distinctive form of evil essentially employed a cynical and distorted form of moral self-justification that he manipulated for his own purposes' (p. 139).

I consider this to be a highly accurate description of evil action or - less metaphysically - of what is going on when we look at actions and practices under non-ideal conditions. However, Stalin's perverted self-justifications were not limited to second-personal address. He was eager to justify himself to others in general, not only to his adversaries. So if such a distortion of morality is possible, then morality cannot be limited to the reasons we give in direct second-personal address.

In order to make room for bad action and bad practices, one must, I think, give up the thesis that autonomy necessarily entails the moral law. Darwall should affirm explicitly that a more moderate conception of practical reason in terms of critical reflection and endorsement can also provide, as Korsgaard demands, a 'bridge into moral territory' (Korsgaard 2007, 20). Critical deliberation is sufficient to convince us of the force of certain moral claims. In my view Darwall is neither committed to a moralizing conception of practical reason or the will, nor is he committed to an account of morality on 'transcendental grounds'.

Various passages in Darwall's text show that he is far from endorsing such strong assumptions. He claims, for example, that the gap from autonomous and critical deliberation to the acceptance of moral principles can be closed by considering our interests, when he writes:

[15] That point is also discussed by R. Jay Wallace who argues against Darwall that agent-neutral reasons might 'turn out to be more weighty or significant at the end of the day, determining what there is most reason for the agent to do' (Wallace 2007: 24).

> 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. (p. 274)

The spirit of that passage shows that Darwall in fact subscribes to a form of contractualism that does not dissociate our agreement on moral principles from our interests as autonomous agents. We can use the notion of 'our interests as autonomous agents' to distinguish between the reasons for inventing the institution of morality and the grounds of normative acceptance. That means: To set up the institution of morality via a contractual agreement is in our interests as autonomous agents, but the recognition of the rules and principles we adopt on such a basis is that they cannot be reasonably rejected. In order to assess what we can reasonably reject the Kantian principles and testing procedures are — a point Darwall affirms — not only helpful but indispensable. But that does not require us to accept the more implausible commitments of Kant's programme.

REFERENCES

Betzler, M. (2009), 'Zweitpersonale Gründe. Was sie sind und was sie uns zeigen', *Deutsche Zeitschrift für Philosophie*, 57: 159-163.

Darwall, S. (2007), 'Reply to Korsgaard, Wallace, and Watson', *Ethics*, 118, 1: 52-69.

------ (2009), 'Eine Antwort auf Monika Betzler, Sebastian Rödl und Peter Schaber', *Deutsche Zeitschrift für Philosophie*, 57: 173-179.

Fichte, J. G. (2000), *Foundations of Natural Right. According to the Principles of the Wissenschschaftslehre.* Edited by F. Neuhouser. Translated by Michael Baur. Cambridge: Cambridge University Press (Cambridge Texts in the History of Philosophy).

----- (2005), *The System of Ethics. According to the Principles of the Wissenschaftslehre*. Translated and edited by D. Breazeale and G. Zöller. Cambridge: Cambridge University Press (Cambridge Texts in the History of Philosophy).

Kant, I. (1996) *The Metaphysics of Morals*. Translated and edited by Mary Gregor with an Introduction by Robert J. Sullivan. Cambridge: Cambridge University Press (Cambridge Texts in the History of Philosophy).

----- (1997) *Critique of Practical Reason*. Translated and edited by Mary Gregor with an Introduction by Andrews Reath. Cambridge: Cambridge University Press (Cambridge Texts in the History of Philosophy).

----- (1998), *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor with an Introduction by Christine M. Korsgaard. Cambridge: Cambridge University Press (Cambridge Texts in the History of Philosophy).

Kleingeld, P. (forthcoming 2010), 'Moral Consciousness and the "Fact of Reason"', in Andrews Reath and Jens Timmerman (eds.) *A Critical Guide to Kant's 'Critique of Practical Reason'*. Cambridge: Cambridge University Press.

Klotz, C. (2002), *Selbstbewusstsein und praktische Identität: Eine Untersuchung über Fichtes Wissenschaftslehre nova methodo*. Frankfurt am Main: Vittorio Klostermann.

Korsgaard, C. M. (1996a), *The Sources of Normativity*. Cambridge: Cambridge University Press.

----- (1996b), *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

----- (1996c), 'Morality as Freedom', in Korsgaard (1996b):159-187.

----- (1996d), 'Kant's Analysis of Obligation', in Korsgaard (1996b): 43-76.

----- (1997), 'The Normativity of Instrumental Reason', in Garrett Cullity and Berys Gaut (eds.) (1997), *Ethics and Practical Reason*. Oxford: Clarendon Press: 215-54. Reprinted in Korsgaard (2008): 27-68.

------ (2002) 'Internalism and the Sources of Normativity.' An Interview. http://www.people.fas.harvard.edu/~korsgaar/CPR.

------ (2007), 'Autonomy and the Second-Person Within: A Commentary on Stephen Darwall's *The Second-Person Standpoint*', *Ethics*, 118, 1: 8-23.

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

------ (2009), *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Nagel, T. (1970), *The Possibility of Altruism*, Princeton: Princeton University Press.

Rödl, S. (2009), 'Darwall gegen Kant: Kant verteidigt', *Deutsche Zeitschrift für Philosophie*, 57: 163-168.

Schaber, P. (2009), 'Achtung vor der Würde von Personen', *Deutsche Zeitschrift für Philosophie*, 57: 169-173.

Tugendhat, E. (1993) *Vorlesungen über Ethik*, Frankfurt am Main: Suhrkamp.

Wallace, R. J. (2007) ‘Reasons, Relations, and Commands: Reflections on Darwall’, *Ethics*, 118: 24-36.

Watson, G. (2007), ‘Morality as Equal Accountability: Comments on Stephen Darwall’s *The Second-Person Standpoint*’, *Ethics*, 118, 1: 37-51.