



Autonomy and the Second Person Within: A commentary on Stephen Darwall's The Second-Person Standpoint.

The Harvard community has made this
article openly available. [Please share](#) how
this access benefits you. Your story matters

Citation	Korsgaard, Christine M. 2007. Autonomy and the second person within: A commentary on Stephen Darwall's The Second-Person Standpoint. Ethics 118(1): 8-23
Published Version	http://dx.doi.org/10.1086/522019
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:3122463
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

SYMPOSIUM ON STEPHEN DARWALL'S
THE SECOND-PERSON STANDPOINT

Autonomy and the Second Person
Within: A Commentary on Stephen
Darwall's *The Second-Person Standpoint**

Christine M. Korsgaard

When you address a claim or a demand to someone, expecting him to respond to that claim as one that gives him a reason for action, you are attempting to issue what Stephen Darwall calls a second-personal reason. Whether you succeed in doing so, according to Darwall, depends on whether certain normative "felicity conditions" of making such a demand are met. Since you expect the other person to take your demand as giving him a reason, you are supposing that you have the authority to bind him by your command. Since you expect him to take your demand as giving him a reason, you are supposing that he can conform to it in the way that one conforms to a reason—by acknowledging its force and imposing it freely on himself, rather than, say, by being frightened into obedience by the fear that you will retaliate if he does not. You are assuming that he has what Darwall calls second-personal competence, the competence to respond to a demand as a reason in the right way. And for both of these reasons, you are also supposing that he is accountable both to you and to himself if he does not conform to the demand that you have made.

Darwall argues that because second-personal reasons have these felicity conditions—the reciprocal authority of both parties, the competence to respond to it, and the accountability that results—the very act of addressing one entails certain recognizably moral commitments. And this is important, because Darwall thinks that these felicity con-

* Stephen Darwall, *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge, MA: Harvard University Press, 2006). References will be inserted into the text using the abbreviation *SPS*.

ditions hold for any attempt to give a genuinely second-personal reason, not just for the issuing of moral demands. In issuing a demand that you expect to be taken as a reason, you commit yourself to the idea that you and the person on whom you make the demand share a normative point of view, from which both of you can see that your demand constitutes a reason for someone in his place to act. Because that is the form of a reason, you are commanding him not merely as, say, a soldier under your military command, but as a person who happens to be a soldier under your military command. To this extent you are committed to viewing him as your equal, and as sharing in authority with you—not the authority you have as his military commander, but the authority of a fellow member of the moral community. For that is the shared normative point of view from which you both can acknowledge that his being a soldier under your military command is a reason for him to obey you. And because you are committed to regarding him as an equal, there are constraints on what you can reasonably command him to do. For if he is to obey your command in the way that one conforms to a reason, by acknowledging its authority and imposing it on himself, then it must be reasonable for him to do that—in particular, it must be consistent with his own authority and dignity as a free and rational being.

Of course this quick summary does not do justice to Darwall's complex conception, but I hope it will give you the general idea. I think that the idea of a second-personal reason is a genuine and important advance in moral philosophy. In particular, it provides a way of articulating what it means to say that you owe something *to* someone, as opposed to just saying that you have a duty to treat him in a certain way. For a duty to treat someone or something in a certain way is not necessarily a duty owed to him, as our duties to care for the environment or artistic treasures for which we happen to be responsible shows. And it also throws light on one important reason why we cannot just pick and choose those whom we are going to count as equals and as the sources of moral claims. If we are to deal second-personally with others in any way, making claims and demands on them and expecting them to take those claims and demands as reasons, then, if Darwall is right, we are committed to seeing those others as our equals in normative authority and so as the sources of moral claims.¹

1. I am tempted to say this is true if we talk to them at all, not as we might chatter to a pet, but with an expectation that they will understand. For given the normativity of meaning, talk is a kind of command—a command to think certain thoughts, although not necessarily to credit them. (For this comparison see Korsgaard, *The Sources of Normativity* [Cambridge: Cambridge University Press, 1996], 136–42.)

Darwall resists the idea that his view shows that slavery involves “conceptual confusion or some sort of pragmatic contradiction” (*SPS*, 265). That's right: it is a normative felicity condition that is violated by addressing a command to a subject that one undertakes to

However, Darwall makes a further claim for the idea of second-personal reasons and the moral commitments that they imply, namely that second-personal reasons in a sense provide the foundation for moral obligation, roughly as understood by Kant. I say "in a sense" because Darwall's argument is modeled on Kant's argument for the foundation of morality in the *Critique of Practical Reason*. And Kant argued there that while freedom is the *ratio essendi* of morality—the reason for its existence—morality is the *ratio cognoscendi* of freedom—the way we know that we are free (5:4n).² As far as I know, Darwall does not explicitly say, mimicking Kant, that "autonomy is the *ratio essendi* of second-personal competence, but second-personal competence is the *ratio cognoscendi* of autonomy," but he seems to have something like that in mind. For he argues that our capacity for autonomous motivation is revealed by—and is revealed only by—our capacity for second-personal transactions of this kind. If that is right, then, strictly speaking, Darwall thinks, like Kant, that autonomy or freedom is the foundation of morality. But unlike Kant, he thinks that we cannot discover our freedom by any reflection that proceeds simply from the first-person standpoint. Darwall therefore opposes his view both to Kant's view, as Darwall himself understands it, and to my own version of Kantianism. These are the arguments that I will address in these comments.

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."³ That may sound paradoxical but it is basically right. What makes this combination of views possible is that I think that the reflective structure of self-consciousness inevitably places us in a relation of authority over ourselves and that we are as a consequence also accountable to ourselves.⁴ By the reflective structure of self-consciousness, I mean the fact that we are conscious of the potential grounds of our beliefs

treat as a mere object. But Darwall should not resist the idea that there is some sort of confusion or contradiction here. If treating a subject as a mere means or an object is the general form of immorality, then there is something especially perverse about both issuing commands to your slaves and lying, for in both of these cases it is the subject's subjectivity itself that is treated as a means.

2. References to Kant's ethical works are inserted into the text using the usual method of citing the volume and page number of the standard German edition of *Kants gesammelte Schriften* (published by Prussian Academy of Sciences in Berlin). The editions of Kant's ethical works I have used are those edited and translated by Mary Gregor for the Cambridge Texts in the History of Philosophy Series: *Groundwork of the Metaphysics of Morals* (Cambridge: Cambridge University Press, 1997) and *Critique of Practical Reason* (Cambridge: Cambridge University Press, 1997). For the *Critique of Pure Reason* I have used Norman Kemp Smith's edition (Boston: St. Martin's Press, Macmillan, 1929), citing the pages of the A and B editions as he does.

3. Both characterizations are on SPS, 11.

4. See Korsgaard, *The Sources of Normativity*, esp. 104–5 and 150–51.

and actions as such. When I am conscious that I am tempted to do something because of something else, I can ask myself whether I should do that, and this amounts to asking whether the consideration on which I propose to act provides a reason. To answer in the affirmative is to adopt a certain maxim of action as a law governing my conduct; to answer in the negative is to command myself to adopt a law of doing the opposite—refraining from the forbidden action and, if necessary, taking positive action to avoid the violation. Thus I act under my own authority as a lawgiver, and I am accountable to myself if I do not. So my reasons—and indeed practical reasons in general—are grounded in the authority the human mind necessarily has over itself. We might put this by saying that because of the reflective structure of human consciousness, I think that every rational agent stands in what Darwall would call a second-personal relation to herself—she has a second-personal voice within.

I am happy to concede to Darwall that this cannot be the whole story. As I mentioned earlier, this does not by itself get us to the idea that I owe a certain kind of treatment to you. Instead it only gets us to the idea that I owe it to myself to treat you in a certain way. We need the second-personal reason for the latter conclusion. But as Darwall himself emphasizes, Kant insists that the second-personal obligation must always be accompanied by an obligation to myself. As Kant says, “I can recognize that I am under obligation to others only insofar as I at the same time put myself under obligation” (6:417). Darwall quotes that remark twice in his book, for it is essential to his argument (*SPS*, 23 n. 47, 248 n. 9). For when I address a demand to you, I know that you can accept the claim that you are under an obligation to me only insofar as you can put yourself under an obligation to yourself to do what I demand. And it is precisely this fact that forces me to concede that you have the same kind of autonomy and dignity that I do and to constrain my demands accordingly.

Since in order to address a second-personal reason to you, and so to obligate you, I have to assume that you have the capacity to obligate yourself, it would be natural to suppose that the capacity to obligate oneself is the prior notion. But this is not Darwall’s view—at least, not so far as epistemic priority, the *ratio cognoscendi*, is concerned. For he thinks it is only in our responsiveness to second-personal reasons that we discover the capacity to obligate ourselves. So the question is why he thinks that and whether he is right.

Darwall mounts his argument on what he takes to be the defects in Kant’s arguments. So examining this question is going to take us into some of the murkiest and most contested parts of the Kantian corpus: the two different arguments that Kant gives for the foundation of morality in the third section of the *Groundwork of the Metaphysics of Morals*

and in the *Critique of Practical Reason* and the question why Kant abandoned the former in favor of the latter. I will try to be brief.

In the *Groundwork*, Kant first establishes that a free will would be an autonomous will, and that an autonomous will would be a moral will—that is, a will under moral law. A free will is one that operates independently of alien causes. This is negative freedom. However since the will is a kind of causality, it must operate in accordance with some law or other. Since the will must operate in accordance with a law, and it cannot be a law alien to itself, it must operate in accordance with its own law—a law it imposes on itself. That makes it autonomous, or positively free. And an autonomous will must be a moral will because the moral law, in its universal law formulation, is simply the law of autonomy: it tells us to act on a law we impose on ourselves (4:446–47). Having established these connections, Kant next argues that we must act “under the idea of freedom” (4:447–48). His point concerns the implications of self-consciousness that I mentioned earlier: we act under our own authority and therefore cannot allow ourselves to be directed by alien causes. Since we act under the idea of freedom, Kant proposes, the laws of the free will, moral laws, apply to us. But then Kant raises a worry: to suppose that we are capable of “taking an interest” in an action—that is, of being motivated to do it for its own sake, because we legislate it to ourselves—just is to suppose that we are autonomous. So Kant worries that the inference from negative freedom to positive freedom does not go through; the argument assumes what it is supposed to prove, namely, that we are autonomous (4:448–50). In the *Groundwork*, Kant replies to the worry with a story about the two standpoints we can take on ourselves and our actions (4:450–54). It will not be necessary for us to go into that part of the story here, as Darwall’s objections do not concern it.

Darwall’s objections to Kant’s arguments center on the idea that if Kant’s argument is to establish the Categorical Imperative as the moral law, Kant must be able to rule out other theories of practical reason, and the conceptions of agency associated with those theories, as somehow incompatible with first-person deliberation.⁵ To show that this is impossible, Darwall sets before us a character he calls the naive practical reasoner (*SPS*, 31, 216–17). The naive practical reasoner desires certain things for his own sake, and he thinks there are reasons to achieve or

5. I do not entirely agree with this. If we could give an otherwise adequate account of practical reason without taking on the commitments of substantive realism about reasons, we would not need to argue against substantive realism over and above doing that. This is because the only reason for accepting the commitments of substantive realism is the worry that we cannot give an adequate account of practical reason without them. I do agree that the Kantian must accept the burden of showing that her account yields a superior conception of agency, however.

promote those things. He does not take his desires themselves to be reasons but rather regards his desires as epistemic states, which provide him with perceptions of the values of objects and states of affairs. He acts on the reasons that he sees as arising from those values: as Darwall calls them, state-of-the-world-regarding, agent-neutral, reasons. Although Darwall rejects this picture of practical reason, he does think that there are reasons of the kind the naive practical reasoner operates with. So Darwall thinks that the conception of agency associated with the naive practical reasoner is perfectly intelligible. But the naive practical reasoner is responsive to reasons in the same way that Darwall thinks a theoretical reasoner is, whereas Darwall thinks that the kind of autonomy that Kant associates with practical reason must amount to more than that. After all, Kant thinks that the autonomous will is an independent source of reasons—or, to put it another way, that the Categorical Imperative generates substantive conclusions about what we ought to do. Whereas the norms of theoretical reason, according to Darwall, are norms of consistency and coherence, which rule out certain combinations of beliefs, but do not rule in any beliefs in particular (*SPS*, 215).⁶

Let us now consider how Darwall turns these considerations against the argument of the *Groundwork*. Kant says that a negatively free will is not directed by alien causes. Darwall replies that if we interpret “alien causes” as causes that might interfere with one’s processes of reasoning, this is true, but it does not establish that the will is an independent source of reasons in the way that is required by Kant’s argument (*SPS*, 224–29). In general, Darwall understands negative freedom as non-interference with the processes of reasoning, and positive freedom as the capacity to follow rational norms. These forms of freedom, he argues, are common to theoretical and practical reason, and do not establish that we must regard ourselves as having autonomous wills in the special sense required by Kant’s conception of practical reason. For having these forms of freedom is compatible with acting on principles other than the Categorical Imperative, or on reasons that are not derived from the Categorical Imperative, like the reasons of the naive practical reasoner.

I think there are three things wrong here. The first is that Darwall is looking for the difference between theoretical and practical reason in the wrong place. For Kant, reason in all of its departments is legis-

6. This cannot strictly be correct, since theoretical reason in Darwall’s sense will rule in beliefs like “X, Y, and Z cannot all be true.” But perhaps these are not beliefs “about the world” in the sense Darwall has in mind.

lative.⁷ The principles of the understanding, being synthetic, are substantive, and they do rule certain beliefs in or out. For instance, the principle that every event has a cause rules in every particular belief of the form “this event must have had a cause” and rules out every particular belief to the contrary.

The second thing I would object to—although obviously this is too long a discussion for the present occasion—is the coherence of the idea of the state-of-the-world-regarding reason. This is because I think that the normativity of reason just is the normativity derived from the mind’s authority over itself—where that includes the authority we have over each other. State-of-the-world-regarding reasons as Darwall conceives them do not seem to get their authority from the mind—rather, they seem to have authority over the mind. Where then does their authority come from? Darwall rightly rejects Moore’s idea that there is simply a way the world “ought to be” because, as he says, “only what can be normatively guided can be subject to oughts” (*SPS*, 286). Darwall proposes instead that we interpret such claims as asserting that certain valuing attitudes toward the state of affairs in question are warranted. For instance, to say that a certain state of affairs “ought to be” is to say that it is desirable or worth bringing about—that desiring it is warranted. Darwall then proposes that, “if we . . . (. . . assume that, as a conceptual matter, being desirable or worth bringing about makes a normative claim on action), then it will follow from the fact that an outcome is valuable that there is some reason for the agent to realize it” (*SPS*, 286). Apparently Darwall’s idea is that bringing in the attitude gives us something that can be normatively guided.⁸ One worry about this move is that

7. Logic for Kant gives laws to thinking, and thinking can be either theoretical or practical, so there is nothing specifically “theoretical” about norms of coherence and consistency. And theoretical reason, according to Kant, also makes its own unconditional demands on the world. It demands that the world be a realm of causally interacting objects, and it also demands that the world conform to the teleological principles of speculative metaphysics: that it be a realm of free agents ordered to the good under the government of a creator who guarantees that ordering. But only the world of appearances meets the first demand, and the world does not, or cannot be known to, meet the teleological demand at all. Practical reason, Kant thinks, provides us with the materials to give positive content to the demands of speculative metaphysics. This fact provides what Kant calls, in the second *Critique*, a “credential” for the moral law (5:48), which he thinks takes the place of the deduction. Practical reason also succeeds in imposing its unconditional demand on one thing in the world, the will: that is the doctrine of the Fact of Reason, discussed in the text below. That, according to Kant, is the difference between practical and theoretical reason.

8. One may also have doubts about whether attitudes can be normatively guided. This is not to deny that attitudes may be, in Scanlon’s sense, judgment sensitive (T. M. Scanlon, *What We Owe to Each Other* [Cambridge, MA: Harvard University Press], 20–22). But the kind of influence reason has on attitudes is clearly quite different from the kind it has on actions: one does not have an attitude “for a reason” or as a result of having

unless seeing something as “worth bringing about” means something other than seeing it as something that there is some reason to bring about, inserting a warranted attitude into the story does not seem to be doing much real work. But another worry is just what this quite different kind of normativity, which is not grounded in the mind’s authority over itself, is supposed to be. Perhaps it is just undue theoretical tidiness that makes me think a unified theory of normativity would be preferable.

The third and most important problem is that Darwall’s conceptions of negative and positive freedom do not do justice to Kant’s argument. Negative freedom is not just a matter of not being derailed by causes interfering with the process of reasoning. I think Kant’s view is that the rational mind is by its very nature under its own authority, and a will that simply accepted the authority of state-of-the-world-regarding reasons *would be* governed by something alien to itself. It would be, in Kant’s words, “consciously receiving direction from any other quarter”—something Kant says that reason cannot do (4:448).⁹ Now Darwall indicates that he agrees with Kant that, quoting Darwall, “rational action involves not just beliefs and desires, but also some norm or principle the agent accepts and implicitly makes her own in acting on it” (*SPS*, 225).

I think Darwall mentions the fact that the agent makes the principle

taken something as a reason for the attitude. Of course we do say things like “you ought not to be afraid.” But when we then go on to spell out the reasons why you ought not to be afraid, we do not expect you to draw the conclusion that you ought not to be afraid and as a result stop being afraid. Rather, we hope the reasons themselves will operate on your fear directly. Your agency is not involved in the same way as it is in acting for a reason.

9. Darwall argues that this idea—that reason must not accept direction from outside—also applies equally to theoretical and practical reason, and so does not explain the distinctive sense in which practical reason is autonomous. As I have already suggested (see n. 7), I think Darwall is looking for the difference in the wrong place. Darwall will want to argue that if we insist on taking negative and positive freedom in a stronger sense than the very weak sense he proposes here, it will seem mysterious that our beliefs have to conform to the evidence of our senses—because that will now seem to violate the principle that reason must not be directed from outside. I think that part of the difficulty is that Kant thinks the achievement of understanding, not the truth of belief, is the enterprise of theoretical reason. And part of the difficulty is that although Kant would agree that we are trying to conceptualize the world that is “there anyway” (as Darwall likes to say), for him that does not mean that our conception is supposed to “match” the world that is there anyway. Of course our beliefs must be true—they are responsible to the world that is there anyway—and this is part of the difference between theoretical and practical reason. But the difference is rather subtle. We can reject an apparent “perception” that does not fit our best scientific understanding of the world, just as we can reject a desire that is ruled out by the Categorical Imperative as not providing any reason to act. But in the first case, we seem to need an explanation of why the mental state in question seemed to be a perception, or why the perception was misleading, whereas we do not seem to need anything further in order to dismiss the outlaw desire, at least for practical purposes.

“her own” because he thinks that is sufficient to relieve her of the charge of being directed from outside. And Darwall also agrees that, “for intelligent pursuit of an outcome I desire to involve my will, I must deliberate on the basis of some rational norm, one I take to apply validly to any possible rational agent. And I must presuppose that I am bound by such norms as a condition of the intelligibility of my own deliberation” (*SPS*, 225). But he argues that this does not require autonomy, because it does not require that the norm in question be the Categorical Imperative: it could just as well be the kind of norm that the naive practical reasoner accepts—a consequentialist norm instructing him to act on state-of-the-world-regarding reasons.

In fact this is a theme that comes up again in Darwall’s criticism of Kant’s argument in the second *Critique*, so before I respond to it, let me take that argument up. By the time he writes the second *Critique*, Kant has decided that “the moral law cannot be proved by any deduction” but also that it does not need a deduction, since it is “firmly established of itself” (5:47). Kant now reverses the argument, arguing from morality to freedom rather than from freedom to morality. The argument begins by establishing that “consciousness of the moral law” is what Kant calls a Fact of Reason. To explain what he means, Kant presents us with a pair of examples. In the first, a man claims that his lust is irresistible if the occasion offers him a chance to satisfy it. We then erect a gallows in front of the house where this opportunity occurs, promise to hang him if he gives way to his lust, and ask him if he can resist it now. Presumably he can. Although there are various ways to take Kant’s point, Darwall supposes—and I agree—that the point is simply that the man’s desire for self-preservation will outweigh his supposedly irresistible lust.¹⁰ In the second example, a man is threatened with the gallows if he refuses to give false testimony against an honorable man whom his prince wishes to destroy on a plausible pretext. Would

10. According to another way to read the story, this first example already shows that the man is free in a sense—he can resist his inclination when prudence demands it. Prudence is one sort of command of reason, and we can follow it, so we must be free. That Kant did hold an idea like this at one point we know, from some remarks he makes in the *Critique of Pure Reason*, in the Canon of Pure Reason. There Kant said that we can know what he calls “practical freedom” through experience, because “the human will is not determined . . . by that . . . which . . . immediately affects the senses; we have the power to overcome the impressions on our faculty of sensuous desire, by calling up representations of what, in a more indirect manner, is useful or injurious” (A802/B830). One might take this example as illustrating that point—the man displays practical freedom in being moved by a representation of the injurious—gratifying his lust will be injurious if he is hanged for it. I do not think this is the right way to read the passage in the second *Critique*, since Kant is clear in the second *Critique* that we learn about our freedom from morality. But it would be possible to argue that Kant intends to be using the pair of examples to be arguing from practical freedom to transcendental freedom.

his love of life—now established as his strongest desire—make it impossible for him to refuse? Kant says that perhaps you would not venture to say whether you would refuse but that anyone “must admit without hesitation that it would be possible for him.” And Kant concludes: “He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him” (5:30).

Darwall accepts the argument, although not Kant’s assertion that it leads to a cognition of freedom. This is because he accepts it for what I think is the wrong reason. He says: “The point seems to follow from the nature of a deliberative (normative) practical judgment. Were he to suppose his desire for self-preservation to be literally irresistible, he would be forced, in reasoning practically in light of that, to conclude he should do something else” (*SPS*, 236). In other words, the person making the judgment cannot decide to do what he knows to be impossible. Darwall goes on to complain that this point does nothing to establish the Categorical Imperative as the exemplar’s principle. A Rossian deontologist, he says, could accept the argument that whatever a person morally ought to do must be something that she can do. But Darwall argues there is no reason for him to accept the Categorical Imperative on that basis or to accept the idea that something like the Categorical Imperative underlies the duties he does accept (*SPS*, 239).

But Kant’s point is not that whatever a person ought to do must be something that she can do. Kant’s point is that in the first person deliberative situation, we know that we can be motivated to do what we believe we ought to do. The “can” here is not, as Darwall says, “that of an open deliberative alternative”; rather, it is something like “can bring myself to.”¹¹ For it is this—the capacity to be motivated by whatever law we give to ourselves, positive freedom—that Kant was worried we might be illicitly presupposing in the “deduction” of morality from freedom in the *Groundwork*. Kant has now decided that we do not have to prove that we can be motivated by the moral law, since each of us in the deliberative perspective knows perfectly well that she can be.

Darwall assumes that the Fact of Reason argument is supposed to lead us to a formulation of the Categorical Imperative. And he wonders how it can do that, since a Rossian deontologist must also agree that he can do what he ought. But the Fact of Reason argument by itself does not have to lead us to a formulation of the Categorical Imperative. For in the second section of the *Groundwork*, Kant has argued that we

11. Kant is using the German version of “ought implies can,” not the English version. In the English version, the fact that I cannot A is used to rule out the claim that I ought to A. In the German version, the fact that I ought to A is used to rule in the claim that I can.

can derive the universal law formulation of the Categorical Imperative analytically from the very idea of a categorical imperative. To put that more clearly, he thinks that the Categorical Imperative can be derived from the very idea of a principle that commands unconditionally. Kant says: "When I think of a categorical imperative I know at once what it contains. For since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary" (4:420–21). What Kant is saying is this: a categorical imperative is a law, to which our maxims must conform. But the reason they must do so cannot be that there is some further condition they must meet or some other law to which they must conform. For instance, suppose someone proposed that one must keep one's promises because it is the will of God that one should do so—the law would then "contain the condition" that our maxims should conform to the will of God. This would yield only a conditional requirement to keep your promises—if you would obey the will of God, then you must keep your promises—whereas the Categorical Imperative must give us an unconditional requirement. Since there can be no such condition, all that remains is that the Categorical Imperative should tell us that our maxims themselves must be laws—that is, that they must be universal, that being the characteristic of laws.

There is a simpler way to make this point. What could make it true that we must keep our promises because it is the will of God? That would be true only if it were true that we must indeed obey the will of God, that is, if "obey the will of God" were itself a categorical imperative. Conditional requirements give rise to a regress; if there are unconditional requirements, we must at some point arrive at principles on which we are required to act, not because we are commanded to do so by some yet higher law, but because they are laws in themselves, laws by virtue of their own nature. The Categorical Imperative, in the most general sense, tells us to act on those principles, principles that are themselves laws.¹² Kant takes it that this means they will be laws by virtue of their form.

Versions of this argument occur in many places in Kant's ethical writings. The argument of *Groundwork* I is a version of this argument, as are the arguments that occur both before and after the presentation of the Fact of Reason example in the *Critique of Practical Reason*. In all

12. The last two paragraphs are lifted from my introduction to Kant's *Groundwork of the Metaphysics of Morals* in the Cambridge Texts in the History of Philosophy edition (Cambridge: Cambridge University Press, 1997).

cases, the basic argument is that an unconditional obligation cannot be characterized by its content alone. Kant believes that when a law is characterized only as having a certain kind of content, what binds us to conform to that law could only be some empirical condition—for instance that we simply like acting in the way the law prescribes, like the sympathetic man of *Groundwork* I, or, supposing that the laws are divine commandments, that we fear divine punishment. In these cases we are not unconditionally bound. An unconditional law must therefore be characterized by universal form rather than by its content: it must be the kind of thing that you find you must will as a universal law.

The importance of that argument for our purposes lies in this: in connection with the *Groundwork* argument, Darwall claims that someone who acts on a consequentialist norm can meet the Kantian requirement that reason should not be directed from outside, so long as she thinks that the consequentialist norm validly applies to any possible rational agent, and so long as she adopts this norm as her own. And in connection with the Fact of Reason argument, Darwall argues that a Rossian deontologist can accept the Fact of Reason argument without thinking that he is committed to the Categorical Imperative, or to the idea that something like the Categorical Imperative underlies the requirements he does accept. But if Kant is right in arguing that the Categorical Imperative can be derived from the very idea of an unconditional requirement, then both of these agents are committed to the Categorical Imperative. For they are committed to thinking that the norms they act on are unconditional rational requirements.¹³ They must therefore think that the principles they propose to follow are the kinds of principles that the Categorical Imperative picks out as unconditionally binding—principles which must be willed as universal laws. Combined with the Fact of Reason, this commits them to being bound by the Categorical Imperative. In other words, Kant's argument, if it works, sets the terms of the debate about which theory of practical reason is correct: the consequentialist and the Rossian deontologist must show that their principles are ones that must be willed as universal laws.¹⁴

I have not said all this just for the sake of disagreeing with Darwall's assessments of Kant's arguments. Rather, I am interested in the effect

13. The Rossian intuitionist is an awkward example for my point, since he accepts a bunch of different norms, each of which is conditional upon the implications of the others in any given case. But presumably the Rossian intuitionist thinks his deliberative conclusion expresses an unconditional requirement, and that is enough for purposes of the argument: he thinks it is a categorical imperative.

14. A parallel move is made later, by John Rawls, when he proposes that any proposed principles of justice must be tested by constructing the original position in which it would be chosen, and seeing whether its assumptions are plausible; see *A Theory of Justice*, 2nd ed. (Cambridge, MA: Harvard University Press, 1999), 102–5, 130–68.

Darwall's objections to Kant's arguments might have on his own arguments. Following Kant, I have been arguing that autonomy, giving laws to ourselves, is an essential feature of the first-person deliberative standpoint. I have suggested that if Kant is right about the nature of action, an agent cannot merely follow some principle other than the Categorical Imperative: rather, she must judge that any principle she follows conforms to the Categorical Imperative. I think Darwall's denial of this creates problems for his own argument.

We can see this by asking whether some of the moves that Darwall makes against Kant can be turned against his own argument. Darwall thinks that an agent can realize both negative and positive freedom while acting in accordance with principles other than the Categorical Imperative, for instance by taking up some consequentialist principle and "making it her own." If so, how does the fact that we hold someone responsible for freely determining herself to respond to our second-personal demands tend to show that we must accept something like the Categorical Imperative? Why aren't negative and positive freedom in Darwall's reductive senses sufficient for the kind of free self-determination involved in second-personal competence? Suppose the commander and the soldier are naive second-personal transagents, who both believe that there is a state-of-the-world-regarding reason for inferiors to obey superiors. Rank, they think, warrants an attitude of submission on the part of the soldier, and this translates into a reason to obey. Darwall may want to argue that state-of-the-world-regarding reasons do not in general entitle us to make demands on each other—that is, in general we cannot demand that people do things just because those things are good. We may reply that what warrants the demand in this particular case is simply the content of the state-of-the-world-regarding reason at hand: the content is that the fact of rank warrants command, submission, and accountability. Nonmoral second-personal demands usually involve established authority relations, and if this kind of move could be made in the case of all second-personal demands except the ones that are already explicitly moral, then how are the presuppositions of second-personal address supposed to provide us with a bridge into moral territory?

The other problem that arises from Darwall's rejection of Kant's arguments concerns the apparently optional character of the second-person standpoint. Suppose someone does not hold herself or anyone else responsible for meeting demands, or resent it when they do not? In chapter 6, Darwall notes that it is a natural response to his argument to agree with him about the presuppositions of second-personal address but then to "deny that second-personal address, so understood, is anything we need have much of a stake in" (*SPS*, 138). Darwall goes on to describe various psychological pressures in favor of addressing and re-

garding people second-personally, citing the fact that even Stalin went in for resentment and rationalization as evidence of this pressure (*SPS*, 138–40). But he concedes that these pressures do not show that there is any incoherence in avoiding second-personal address. And he also concedes that citing the psychological difficulty of avoiding resentment and other such phenomena does not quite answer “the philosophical question . . . whether we should see things in this way” (*SPS*, 278). But if we do not have to enter the second-person standpoint, why must we have second-personal reasons at all?

At the beginning of chapter 11, Darwall appears to be about to address this problem. He says: “Even if taking up the second-person stance commits us to equal dignity and autonomy, that is consistent with it and its associated commitments being no more than rationally optional” (*SPS*, 277). And he says that he is going to respond to this worry by arguing that “the second person standpoint gives us a perspective on our own agency that enables us to appreciate a fundamental difference between theoretical and practical reason and so improves our grasp of reasons for acting” (*SPS*, 277).

But, we may wonder, does someone who does not engage in second-personal address fail to see reasons that she has, or only reasons that she might have had if she had engaged in second-personal address? Darwall certainly suggests that she fails to see reasons that she has. He says: “Were someone somehow to avoid taking [the second-person standpoint] up . . . she would fail to appreciate what we, who have taken it up, can validate as reasons from a more comprehensive view that includes it” (*SPS*, 277).¹⁵

And he also says elsewhere that “not even universal disrespect can destroy the dignity of free and rational persons. . . . Equal dignity is nothing anyone can bestow, so neither is it anything any person can

15. The formulation here, which suggests that second-personal reasons are real because we can validate them by viewing them from a more objective standpoint, recalls Thomas Nagel’s views about the role of assuming more objective standpoints in *The View from Nowhere* (New York: Oxford University Press, 1986). I think this is no accident. In fact, I think we can usefully see Darwall’s book in general as presenting us with a variant of Nagel’s argument in his first book, *The Possibility of Altruism* (Princeton, NJ: Princeton University Press, 1970). Nagel argued that we can capture the essence of moral reasons if we think about how a certain familiar form of moral argument is supposed to work. When we ask “how would you like it if someone did that to you?” Nagel proposed, we invite the agent to view himself and his victim as interchangeably situated “someones,” and to conclude that doing that to any “someone” is objectionable. In a similar way, Darwall’s argument invites us to view ourselves and others as interchangeably situated “yous” and to conclude that doing certain things to anyone with whom we stand in a “I-thou” relationship is objectionable. Although Darwall doesn’t make this comparison explicitly, he does draw a contrast between his own view and Nagel’s along these lines at *SPS*, 102.

remove through disrespect" (*SPS*, 144). So presumably Darwall does think that anyone who failed to take up the second-person standpoint would fail to see reasons that they have anyway, reasons for according dignity to every rational being.

In the body of the chapter 11 discussion, Darwall distinguishes the formal aims of action and belief from their substantive aims. The formal aim of belief is to believe whatever there is reason to believe; the substantive aim is to believe what is true. The formal aim of action is to do whatever there is reason to do. The naive practical reasoner, Darwall goes on to argue, supposes that the substantive aim of action is to bring about valuable states of the world. If that were true, Darwall claims, "it would be psychologically impossible to decide, intend, or do something on any other grounds" (*SPS*, 287). When we grasp the existence of second-personal reasons, Darwall says that "we see that what seemed to be true from a naive practical standpoint, that action has the substantive aim of bringing about valuable outcomes, is an illusion" (*SPS*, 290). What then is the substantive aim of action, according to Darwall? Oddly, the question comes up again only in a footnote to a passage in which Darwall is explaining why second-personal reasons can fit into a constructivist framework. Though I am unable to reconstruct Darwall's reasoning here, he canvasses two conclusions, that action has no substantive aim, or that the substantive aim of action is autonomy (*SPS*, 295 n. 30). But if the latter is true, how is the naive practical reasoner psychologically possible?¹⁶ According to Darwall, he does not aim at autonomy when he acts on his naive reasons. And if part of the lesson we learn when we take up the second-person standpoint is that the naive practical reasoner is not possible after all, how can the claim that he is possible be used in the argument against Kant?

I do not wish to rest too much weight on that particular argument, which, after all, only appears in a footnote. My worry here can be put in more general terms. On Darwall's account, the second-person standpoint does not seem to be unavoidable, the way the standpoint of first-personal deliberation is. If we are autonomous, then the person who declines to take up the second-person standpoint fails to know something intimate and important about his own agency. But unless he has a reason to take up the second-person standpoint and its presuppositions, it is possible that he will never know. And that conclusion can be generalized: if it were not for the fact that it is so psychologically difficult to avoid holding one another accountable, it seems as if we might never

16. Kant's argument that anyone who accepts an unconditional requirement is in effect accepting the Categorical Imperative amounts to an argument that it is psychologically impossible to respond to an unconditional demand without responding to the Categorical Imperative.

have discovered that we are autonomous. But then how could we ever have been autonomous? And how do we go about “presupposing” the autonomy of others in second-personal address, if we have no independent access to this feature of human agency?

On Kant’s account as I understand it, by contrast, the second-person standpoint is unavoidable, because I do not have to discover, by making and responding to demands on others, that I am answerable to myself. That fact is made clear to me by the voice of the second person within.

Copyright of Ethics is the property of University of Chicago Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.