

Moral Obligation, Accountability, and Second-Personal Reasons

MICHAEL SMITH

Princeton University

JADA TWEDT STRABBING

Princeton University

Stephen Darwall's aim in *The Second-Person Standpoint* is to reorient the way we think about moral obligation and accountability.¹ Though he advances on many fronts, we focus on what we take to be the main lines of argument.

Darwall begins by discussing second-personal reasons, a sub-class of agent-relative reasons, where the defining feature of such reasons is their connection with accountability. What makes it appropriate for others to hold us accountable for failing to act in a certain way, according to Darwall, is that we have sufficient second-personal reasons for acting in that way. However, he claims, there is also a conceptual connection between moral obligation and accountability, so moral obligation too must be second-personal. We are morally obliged to act in a certain way only if we have sufficient second-personal reasons for so acting. He concludes that accounts of moral obligations that see them as grounded in agent-neutral reasons, and those which hold that they may be overridden by non-moral obligations, are both radically mistaken.

We wish to explain why, as we see it, Darwall's own arguments are inconclusive. For all he says, moral obligations may be grounded in agent-neutral reasons alone and may themselves be overridden by non-moral obligations.

¹ Stephen Darwall, *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge: Harvard University Press, 2006). All page references are to this work.

1. The Characterization of Second-Personal Reasons

Darwall begins by discussing an example involving both second-personal reasons and agent-neutral reasons. Since he constantly returns to this example, we assume that it reveals the crucial features of such reasons.

Suppose someone steps on your foot, causing you pain. One reason he has for removing his foot is agent-neutral: pain is a bad thing, so removing his foot would stop something bad from happening. When you give someone a reason of this kind, Darwall says,

...you wouldn't so much be addressing it to him as getting him to see that it is there anyway, independently of your getting him to see it or even of your ability to do so. (6)

The agent-neutral reason thus requires nothing from you by way of a capacity to recognize the badness of your pain; it requires nothing from you by way of authority to demand that the agent remove his foot; and it requires nothing from him by way of such capacities either. These are, however, exactly what's required for the second-personal reason. In bringing this reason to an agent's attention, Darwall suggests,

You might say something that asserts or implies your authority to claim or demand that he move his foot and that simultaneously expresses this demand. You might demand this as the person whose foot he is stepping on, or as a member of the moral community, whose members understand themselves as demanding that people not step on one another's feet, or both... The reason would not be addressed to him as someone who is simply in a position to alter the regrettable state of someone's pain or of someone's causing another pain... It would be addressed to him, rather, as the person causing gratuitous pain to another person, something we normally assume we have the authority to demand that persons not do to one another. (7)

This second-personal reason is one that someone can sensibly accept, Darwall tells us,

...only if he also accepts your authority to demand this of him (second-personally). And if he accepts that you can demand that he move his foot, he must also accept that you will have grounds for complaint or some other form of accountability-seeking response if he doesn't. Unlike the first state- or outcome-based reason, this...is second-personal in the sense that, although the first is conceptually independent of the second-personal address involved in making claims and holding persons responsible, the second is not... As second-personal reasons always derive from agents' relations to one another, they are invariably fundamentally agent-relative. (8)

Thus it seems that someone has a second-personal reason for getting off your foot in virtue of having an agent-neutral reason that meets the following conditions. First, his stepping on your foot must be the cause of your pain; second, you must be the sort of person who can recognize the fact that your pain provides him with an agent-neutral reason not to step on your foot; third, you must have the authority to demand that he not do so; and fourth, he must have the capacity to recognize and respond to the three conditions just mentioned. When someone has an agent-neutral reason for getting off your foot that stands in these relations to himself and to you he additionally has an agent-relative reason for getting off your foot that is *second-personal*, a reason that derives from your authority to demand of people that they not cause bad things to happen to you.² So Darwall seems to suggest.³

However, if this is right, shouldn't we suppose that there is a further distinctive class of agent-relative reasons deriving from facts about what we *allow*? Imagine someone who allows two others to cause gratuitous pain when he could prevent their doing so by stepping on your foot so causing you pain. Suppose that both you and those in pain have the capacity to recognize that there is agent-neutral reason not to allow this; that the agent has the capacity to recognize and respond to these facts; and that you, and those in pain, are in a position to draw the existence of the agent-neutral reason meeting these conditions to his attention and have the authority to do so. In failing to step on your foot, he thus fails to exercise his capacity to recognize and respond to an agent-neutral reason that meets conditions like those Darwall says are met by second-personal reasons. The only difference is that they concern what is *allowed*, not what is *caused*.

The question is why agent-neutral reasons meeting these conditions don't form a distinctive class of, say, second-personal* reasons, reasons that would presumably derive from your authority to demand that people not allow bad things to happen. This question is urgent because, if such reasons exist, the normative significance of second-personal reasons plausibly derives from what they have in common with second-personal* reasons. The number of dimensions along which

² Why does the second-personal reason require that all of these conditions are met, rather than some subset? Suppose all of the conditions are met, except that no evidence of the whereabouts of your foot is available to the person causing you pain. Does a third party who makes that evidence available thereby make available a second-personal reason to get off your foot that existed all along, or does he create the reason? We are not sure.

³ We put the point cautiously because this involves a generalization from Darwall's stepping-on-your-foot example. Does he think that all second-personal reasons have the features possessed by the second-personal reason in this example? If not, our criticisms concern those that do possess those features.

second-personal reasons differ in normatively significant ways from agent-neutral reasons would thus be significantly reduced. We assume that Darwall will respond that they do not exist because, though we have the authority to demand that others not cause us pain, we lack the authority to demand that they do not allow pain to occur. But we see two problems with this response.

First, the normative significance of second-personal reasons would then depend on the normative significance of the doing/allowing distinction. Such doubts as we have about this distinction would thus become doubts about second-personal reasons. Second, we cannot see why we lack the authority to demand that people not allow pain to occur. Consider once again the scenario. The mere fact that someone allows pain to occur, when he has the option of not allowing it to occur, may well be insufficient to give others the authority to demand of him that he not allow it. He may lack the capacity to recognize and respond to the agent-neutral reason he has not to allow it; or the others may lack the capacity to recognize the agent-neutral reason he has not to allow it; or he may lack the capacity to recognize their capacity to recognize the agent-neutral reason he has not to allow it; etc. But if these further conditions are met, then it seems to us that they have all the authority they need to make such a demand. He has the agent-neutral reason, after all, and he and they have the capacity to recognize and respond to that fact, so why aren't they entitled to point this out to him?

2. Moral Obligation, Second-Personal Reasons, and Excuses

So far we have seen that second-personal reasons concern what agents cause but not what they allow. We now want to focus on other features of such reasons.

Darwall forges a link between moral obligation and second-personal reasons by pointing out the conceptual connection between moral obligation and accountability.

We hesitate to impute wrongdoing unless we take ourselves to be in the range of the culpable, that is, unless the action is such that the agent is aptly blamed or the object of some other form of accountability-seeking reactive attitude if she lacks an adequate excuse. (93)

Second-personal reasons secure the connection with accountability because, when someone has such a reason, others are guaranteed to have the authority to make demands of him. Moral obligations are thus supposed to entail corresponding second-personal reasons.

There are, however, two ways in which excuses might work. Suppose someone steps on your foot, causing you pain, but then offers, as his excuse, that he couldn't have known of your foot's whereabouts. Does his excuse show that, though he acted wrongly, it would be inappropriate to hold him accountable? Or does it show that he didn't really act wrongly? If it works in the second way, the excuse reveals that what's really wrong is for someone to step on your foot, so causing you pain, *when he could have known that that's what he was doing*. If the excuse works in the first way, by contrast, the obligation is left intact by the inability to know: the agent isn't at fault, even though he acted wrongly.

This question about how excuses work is important because second-personal reasons ensure that the features that constitute excuses for wrongdoing, when we think of excuses in the first way, are absent. People do not have a second-personal reason not to cause others pain when they cannot know that what they are doing would cause them pain; they cannot have a second-personal reason when they are unable to respond to their recognition of that fact; and so on. Excuses for failures to act on second-personal reasons thus seem to function in the second way, not the first. They clarify whether second-personal reasons are really present.

However, as Darwall acknowledges, many theorists with similar views to his about the connection between moral obligation and accountability think that excuses function in the first way. It is, he tells us,

...consistent with these views that there remains a distinction between the wrongness of acts and the blameworthiness of agents. Someone may not be to blame for wrongdoing if he has an excuse. (93)

But if, as these theorists think, people can act wrongly and yet have an excuse for so acting, but people cannot have a second-personal reason to act in some way and yet have an excuse for not acting on that reason, then it follows that, contrary to Darwall, moral obligations cannot require corresponding second-personal reasons.

Putting this observation about the way in which excuses function together with the conclusion of the first section, we get the following result. The additional conditions Darwall insists must be present for an agent who has an agent-neutral reason to get off your foot to have, in addition, a second-personal (or second-personal*) reason to get off your foot are simply those whose presence ensure that the agent has no excuse for failing to act on the agent-neutral reason. But since an agent can act wrongly and yet be excused, the upshot is that, for all Darwall says, the agent's moral obligation may be supported entirely by the

agent-neutral reason. He may act wrongly simply in virtue of stepping on your foot, thereby failing to act on an agent-neutral reason not to cause or allow pain. The additional conditions that are required for him to have a second-personal (or second-personal*) reason not to step on your foot simply ensure that his stepping on your foot is both wrong and inexcusable.

3. Moral Obligation, Accountability, and Sufficient Reasons

Finally, we turn to Darwall's claim that moral obligations must be supported by sufficient reasons. His argument for this claim is that, since it would only be appropriate for others to hold us accountable for failing to act in a certain way if our acting in that way were something we have sufficient reason to do, it follows that what we are morally obliged to do must itself be something that we have sufficient reason to do.

[I]f someone were able to establish that she did in fact have good and sufficient reason for a putative violation of a moral obligation, then it seems she has accounted or answered for herself and shown she did no wrong. When we charge her with wrongdoing, therefore, we must be implying that she can not provide such an account. (98)

We see the appeal of his account of wrongdoing, but we are unsure how to integrate it with a plausible account of moral permissibility and moral obligation.

Darwall's idea seems to be that if an agent shows that she has sufficient reason to act as she did, she accounts for herself and does no wrong, and hence was not subject to a demand to act otherwise. But there are two different things that might be shown. One is that she wasn't subject to a demand to act otherwise because, though she was subject to various demands, what she did accorded with them. The other is that she wasn't subject to any demands in the first place. We assume that Darwall has the first in mind, not the second: you don't have a blanket moral permission to do what you like simply in virtue of the fact that, because you lack the capacities you must have in order to be subject to demands, you are not subject to any demands at all. We therefore take it that an agent's action is morally permissible just in case, though she was subject to various demands, what she did accorded with them. Let's test the plausibility of this idea.⁴

⁴ We adapt an example of Susan Wolf's from "Above and Below the Line of Duty", *Philosophical Topics* (14) 1986, 131-148.

Imagine a professor who promises to meet a student. When the time comes, however, she discovers that the meeting conflicts with two lectures she's curious to attend: Gibbard is giving a lecture, someone whose work she admires but hasn't heard speak, and, at the same time, so is Railton, someone whose work she also admires but whose lectures she has attended before. The professor therefore faces a choice between: acting on her moral reason to keep her promise; acting on her non-moral reason to attend Gibbard's lecture; and acting on her non-moral reason to attend Railton's lecture. Imagine that the student's reason for scheduling the appointment was sufficiently trivial that the non-moral reason she has to attend Gibbard's lecture outweighs the moral reason to keep her promise, but was sufficiently weighty to outweigh the non-moral reason she has to attend Railton's lecture. What is the professor morally permitted to do? As we have seen, this turns on what demands she is subject to.

Darwall is plainly committed to the conclusion that the professor does no wrong if she breaks her promise and attends Gibbard's lecture, for he thinks that she can account for herself by pointing out that she has sufficient reason for so acting. Attending Gibbard's lecture must therefore accord with the demands to which she is subject. What if she keeps her promise, notwithstanding the more weighty non-moral reason to attend Gibbard's lecture? Would she act wrongly? We think not. Keeping her promise must therefore also accord with the demands to which she is subject. What if she attends Railton's lecture? We assume that this would be morally impermissible and hence would not accord with the demands to which she is subject. But if keeping her promise and breaking her promise by attending Gibbard's lecture are both morally permissible, but attending Railton's isn't, then we must suppose that she is subject to the demand that she either keep her promise or attend Gibbard's lecture.

The crucial question, however, is whether a plausible rationale can be given for this demand, and we see no clear way in which it can. Consider the student's point of view. It is obscure to us why he would demand that she either keeps her promise or attends Gibbard's lecture, as it is not his business to demand anything about the lecture. The professor will, of course, be rationally justified in the student's eyes if she goes to Gibbard's lecture, but not if she goes to Railton's. However it does not seem that he can *hold her* to going to the Gibbard lecture if she breaks her promise. He can only hold her to the promise and then decide whether she has a good enough reason to be excused for the violation. Indeed, not only does the student not make this disjunctive demand, no one else seems to make it either. Keeping her promise thus appears to be the only demand in play. The professor therefore acts in

a way that is morally impermissible, albeit on sufficient reasons, if she attends Gibbard's lecture.

Perhaps Darwall can respond to this objection. But, even if he can, we think that there is another problem with his account of the relationship between moral obligation, accountability, and sufficient reasons. Is it true, as he seems to imply, that we *fully* account for ourselves if we establish that we have good and sufficient reason for a putative violation of a moral obligation? We think not. Moreover, we think that he himself explains what else we might need to establish, providing the seeds of an alternative account of what is required for accountability in which sufficient reasons are only part of the story.

Elsewhere Darwall admits that, even when we do what we are morally obliged to do, we can still be held accountable if we do so for the wrong kind of reason.

[T]here may be cases where we take ourselves to have the authority to blame someone and make demands of them even when we don't think they act wrongly, as for example, when someone does the right thing but for utterly unconscionable reasons, say, out of a virulent racism. (93, fn3)

This suggests a rather different picture of what we can be held accountable for, a picture according to which accountability concerns not so much what we do as what we are motivated to do.

To see this, think once again about the professor. Does the mere fact that she acts on her more weighty non-moral reason to attend Gibbard's lecture show that she *fully* accounts for herself? We think not. There are two possibilities to consider. In one, though she acts on the better non-moral reason, she feels some regret for having failed to act on the moral reason, where the regret can be seen as the residual psychological effect of her overridden motivation to keep her promise. In the other, though she acts on the better non-moral reason, she feels no regret at all for having failed to act on the moral reason: she wasn't motivated to keep her promise in the first place. We think that in the second case, by contrast with the first, the professor plainly fails to fully account for herself and that her failure is completely analogous to the failure in the case Darwall considers of someone who is motivated by virulent racism.

What these failures have in common is that the agents in question are motivated to act in ways where those motivations are not themselves supported by the moral reasons in play. Correspondingly, it seems to us that we also demand of agents that the strength of their moral motivations, both their overriding motivations and their overridden motivations, tracks the importance of the moral reasons in

play, both the sufficient moral reasons and the pro tanto moral reasons that are insufficient. The upshot is that there isn't the tight connection between accountability and sufficient reasons that Darwall imagines. Accountability has a much broader purview.⁵

⁵ We gratefully acknowledge helpful conversations with Stephen Darwall, Gary Watson, and the students and faculty who attended MS's graduate seminar on *The Second-Person Standpoint* at Princeton, Spring 2006.