1

DEONTOLOGICAL MORAL OBLIGATIONS AND NON-WELFARIST AGENT-RELATIVE VALUES

Michael Smith

Abstract

Many claim that a plausible moral theory would have to include a principle of beneficence, a principle telling us to produce goods that are both welfarist and agent-neutral. But when we think carefully about the necessary connection between moral obligations and reasons for action, we see that agents have two reasons for action, and two moral obligations: they must not interfere with any agent's exercise of his rational capacities and they must do what they can to make sure that agents have rational capacities to exercise. According to this distinctively deontological view of morality, though we are obliged to produce goods, the goods in question are non-welfarist and agent-relative. The value of welfare thus turns out to be, at best, instrumental.

Many theorists think that two related claims will occupy centre stage in any plausible moral theory. The first is that we should bring about more rather than less of what's of intrinsic value. The second is that welfare has intrinsic value. Putting these two claims together, they suppose that any plausible moral theory will tell us that we should produce more rather than less welfare.

Richard Arneson, who is typical of the theorists I have in mind, puts the point this way:

> The concept of intrinsic value is not merely a building block in consequentialist theories, and if this concept (or the best revision of it we can construct) is found wanting, the loss would have wide reverberations. More is at stake than the status of consequentialism. I suspect any plausible nonconsequentialist morality would include as a component a principle of beneficence. In a consequentialist theory some beneficence principle is the sole fundamental principle; in a nonconsequentialist theory beneficence would be one principle among others. Whatever its exact contours, a beneficence principle to fill its role must rank some states of the world as better or worse, and direct us to bring about the better ones within the limits

Developing Deontology, First Edition. Edited by Brad Hooker.

> imposed by other principles that introduce moral constraints and moral options. We need some commensurability, a measurable notion of good. We need the idea of what is good simpliciter. (Arneson 2010, p. 741)

As Arneson's remarks make plain, the idea isn't just that we have obligations to bring about states of the world that track the amount of welfare in those states of the world. The idea is that we have obligations to bring about states of the world that track the amount of welfare in those states of the world *independently of whose welfare it is*. I take it that this is what Arneson is getting at with his talk of "intrinsic value" and "good simpiciter". The relevant contrast here is with the idea that our obligations track the amount of welfare we bring about, but that the welfare in question is (say) our own welfare, or the welfare of ourselves and our loved ones, or that of our community. To be more precise, then, Arneson thinks that any plausible moral theory will acknowledge that welfare has agent-neutral value, not merely agent-relative value, and that the agent-neutral value of welfare gives rise to an obligation to produce more rather than less welfare independently of whose welfare it is (Smith 2003).

Though I am sympathetic to the idea that any plausible moral theory will tell us that we should produce states of affairs with more rather than less value (Sen 1982; Sen 1988; Broome 1991; Dreier 1993; Smith 2003; Smith 2009), I think that the most plausible such theories will tell us that the values in question are *non-welfarist* and *agent-relative*. My disagreement with Arneson is therefore just about as complete as it could be. In what follows I want briefly to explain why I think that this is so. Though my reasons are somewhat abstract and require a significant detour in order to be spelled out, the substantive moral view which they lead me to embrace should sound familiar, as it is just a standard deontological view of the nature of our moral obligations, albeit one grounded in the existence of agent-relative values. Unfamiliar though my reasons might be, my hope is thus that they will suffice to show how and why a plausible moral theory might eschew both agent-neutral values and the obligations to which they would give rise.

1. A Familiar Puzzle

Whatever else it does, a moral theory will tell us what our moral obligations are. Since if we have a moral obligation to act in a

certain way, it follows that we have a reason to act in that way, this entails that a moral theory will tell us what some of our reasons for action are (compare Darwall 1983, Darwall 2006). But since if we have a reason to act in a certain way, it follows that we would desire that we act in that way if we were to deliberate correctly, this entails that a moral theory will tell us which desires we would have if we were to deliberate correctly (compare Williams 1981 and Scanlon 1998 Appendix). This gives rise to a familiar puzzle.

Hume famously argues that there is a difference between the way in which reasons relate to beliefs, on the one hand, and desires, on the other (Hume 1740). Beliefs are states that purport to represent things as being the way they are. They are therefore "judgement-sensitive attitudes", to use Scanlon's term, because it follows that they are sensitive to the reasons that bear on the truth of the proposition believed (Scanlon 1998, ch. 1). Desires are, however, different. For while certain desires are sensitive to reasons that bear on the truth of propositions believed, these desires are all *extrinsic*. *Intrinsic* desires, by contrast, are not the sort of psychological state that could be sensitive to reasons. Hume thus holds that while extrinsic desires are judgement-sensitive attitudes, intrinsic desires are not.

By way of illustration, suppose I desire to have a pleasurable experience and believe that, since it would be pleasurable to eat a peach, the way for me to have a pleasurable experience is to eat a peach. Assuming that I am instrumentally rational, I will form an extrinsic desire to eat a peach, where this extrinsic desire is an amalgam of my desire to have a pleasurable experience and my belief about the effects of eating a peach on my pleasure: the extrinsic desire is just these two states hooked up in a state of readiness to cause my eating a peach if I don't want to do something else more (Smith 2004). Being an amalgam of desire and belief, my extrinsic desire is sensitive to reasons that bear upon the truth of the proposition that expresses the content of its belief component: that is, it is sensitive to reasons that bear on whether my eating a peach would be pleasurable. But the desire to have a pleasurable experience itself, assuming that it is not an amalgam of some further desire and belief – in other words, assuming that it is an intrinsic desire – is not sensitive to any reasons that bear upon the truth of anything. Though a sensitivity to reasons may change our extrinsic desires, it will do nothing to change our intrinsic desires, as no reasons have any impact on them. So, at any rate, Hume argues.

This is bad news, if it is true. For when we put this together with the claims about moral obligation and reasons for action described earlier, we derive a contradiction. Imagine a husband who has an obligation to treat his wife better than he does, but whose intrinsic desires are thoroughly nasty. Given that he has an obligation to treat his wife better than he does, it follows that he has a corresponding reason for action, and given that he has the corresponding reason for action, it follows that he would desire to treat his wife better than he does if he were to deliberate correctly. But given what Hume tells us about the relationship between reasons and intrinsic desires, it follows that his deliberating correctly would do nothing to change his intrinsic desires. No reasons to which he could be sensitive would have any impact on them. So long as the man we are imagining has made no deliberative error in deliberating *from* his intrinsic desires, it therefore follows that he would not desire to treat his wife better if he deliberated correctly. Contradiction.

2. Rethinking Hume's Strictures

To avoid this conclusion, we have to rethink Hume's strictures on the relationship between reasons and intrinsic desires. The assumption we have been making so far is that agents can deliberate correctly independently of the intrinsic desires that they happen to have. But, as I will now argue, it turns out that this assumption is false. In order to deliberate correctly, rational agents must have certain intrinsic desires (see also Smith 2010).

Imagine someone who believes for reasons by inferring that q from two premises: the premise that p and the premise that p implies q. Now imagine the moment at which he remembers having settled that p is the case, he is in the process of settling that p implies q, and he anticipates the possibility of going on to perform the inference and form the belief that q. What would an agent who has and exercises the capacity to believe that q for reasons have to be like at that very moment? Trivially, he would have to have and exercise the capacity to believe for reasons, as he is in the process of settling that p implies q. But are there any other psychological states that he would have to have or lack? It seems that there are. He would have to be able to rely on his past self who settled that p; he would have to be able to be vigilant at this very moment in settling that p implies q; and he would have

to be able to rely on his future self to draw the inference that q. The mental attitudes constitutive of reliance and vigilance are thus both required.

Let's start by asking what would be required for our agent to be vigilant at this very moment. One way in which an agent may fail to believe for reasons, even when he has the capacity to believe for reasons, is by having an effective desire to believe something whether or not that thing is supported by reasons. This is what happens in wishful thinking. An agent who has and exercises the capacity to believe for reasons – our imagined agent who is in the process of settling that p implies q, for example – must therefore lack such an effective desire. Note, however, that there is more than one way in which he might lack such a desire. One would be to lack any desire at all to believe certain things rather than others, whether or not they are supported by reasons. But given that believing certain things rather than others, whether supported by reasons or not, is something that will contribute to the satisfaction of intrinsic desires that nearly everyone has, this is unrealistic. Even the ubiquitous desire to have pleasurable experiences tells in favour of acquiring certain beliefs, independently of whether they are supported by reasons. Think of the pleasure you derive from believing that your partner is faithful to you. If a fully rational agent is robustly to possess and fully exercise his capacity to believe for reasons, then he must have the wherewithal to cope with the potential deleterious effects of having ordinary desires like the desire to have pleasurable experiences.

Another, and much more realistic, way in which fully rational agents could be vigilant at this very moment, given that they may well have ordinary desires that augur in favour of their having certain beliefs rather than others, is thus by having a stronger desire not to allow those desires to be effective. Suppose, for example, that a fully rational agent desires to have pleasurable experiences, and suppose that this leads him to desire to believe that his partner is faithful to him, independently of the reasons. If he has a much stronger desire not to allow his exercise of his capacity to believe for reasons to be undermined, then the potentially deleterious effects of his desire to have pleasurable experiences would be mitigated. It would lead him to monitor himself to make sure that he isn't being led astray by that desire. Any agent, if he is robustly to possess and fully exercise the capacity to believe for reasons, must therefore have such a desire. Only so could he be on guard against the permanent possibility of engaging in wishful thinking.

Now let's focus on what's required for a fully rational agent to be able to rely on his past and future self. To be able to rely on his future self to draw the inference and believe that q, once he has settled that p implies q, an agent must similarly be on guard against the possibility of his presently having effective desires to undermine his future self's exercise of his capacity to believe for reasons. The way in which to do this, given the psychological resources available to him, is by having another desire like the one already posited, but with a slightly different content. For much the same reasons as before, then, an agent who robustly possesses and fully exercises the capacity to believe for reasons will have to desire that he does not undermine his own future exercises of his capacity to believe for reasons.

What if a fully rational agent foresees that he will be unable to play his role in the exercise of his capacity to believe for reasons in the future? Imagine, for example, that he is involved in a complex chain of reasoning and he foresees that, at some future stage, he will have a diminished capacity to believe for reasons. Perhaps he anticipates that he will have a debilitating headache, and he now has available a pill which, if taken later, would remove the headache. If he is robustly to possess and fully exercise his capacity to believe for reasons in such circumstances, then in order to be able to rely on his future self to play his part, it wouldn't be enough for him to desire not to undermine his future exercise of his capacity to believe for reasons. He would have to desire, more positively, that he now does what he can to help his future self have the required capacities so that he can play his part. In our example, the agent would have to desire to hold on to that pill so that his future self could take it. Absent such a desire, he could not rely on his future self to play his part.

Now consider an agent's past self. The agent we have been imagining, who robustly possesses and fully exercises the capacity to believe for reasons and is in the process of settling that p implies q, also has to be able to rely on his past self having settled that p for reasons, and not having had effective desires to interfere with his current exercise of his capacity to believe for reasons. This last is essential because only so will he be entitled to believe that he is not currently in the grip of an illusion, planted by his past self. In doing this, he seems to count not just on his past self's having possessed and exercised the capacity to believe for reasons, but also having had the very same standby desires as he currently has to ensure that he doesn't engage in wishful thinking

or interference with his present or future self's exercise of his capacity. For so long as his past self had those same desires, his past self will indeed have exercised his capacity to believe for reasons and will not have interfered with his current self's exercise of his capacity to believe for reasons; he will not have planted an illusion.

Relatedly, if an agent is robustly to possess and fully exercise the capacity to believe for reasons, then he also has to be able to rely on the non-interference of other rational agents, assuming that there are such agents. He has to be entitled to believe that he isn't currently in the grip of illusion planted by them. It seems to me that in the special case in which the agent is a member of a community of fully rational agents, this too is grounded in the reasonableness of his supposing that all rational agents, if they are robustly to have and fully exercise their own capacities to believe for reasons, must desire not to interfere with other rational agents exercises of their capacities. For to suppose that rational agents do not extend their concern for non-interference to other rational agents in this way is to imagine that they make an arbitrary distinction between certain of those on whom they must rely – that they make a distinction between their reliance on themselves and their reliance on others – despite the fact that all of those on whom they must rely, insofar as they exercise their capacity to believe for reasons, have the very same interest in the non-interference of others as they have themselves. Since a rational agent would make no such arbitrary distinction, I take that it that his concern not to interfere extends to other rational agents as well.

A similar line of thought suggests that the more positive desire that we earlier saw a fully rational agent would have to have in order to rely on his future self, the positive desire to do what he can to help ensure that his future self has the capacity to believe for reasons, is also an instance of a more general desire that extends to rational agents as such. Fully rational agents must desire to do what they can to help rational agents as such have the required capacities to do their part in what's required for them to believe for reasons. Again, to suppose that rational agents do not extend the desire that they do what they can to ensure that they themselves have rational capacities in the future to other rational agents is to imagine that they make an arbitrary distinction between certain of those on whom they have to rely. To repeat, rational agents would make no such arbitrary distinction.

I said earlier that nothing in the argument turns on the initial focus on what's required to robustly exercise the capacity to believe for reasons. We could just as easily have made all of the same points by asking which desires an agent must possess if he is robustly to possess and fully exercise the capacity to be instrumentally rational: that is, if he is robustly to possess and fully exercise his capacity to form extrinsic desires in the light of his background intrinsic desires and his beliefs about what's required for their satisfaction. For in this case too, it seems that a fully rational agent could robustly possess and fully exercise his capacity only if he is on guard against the possibility of his own interference with his exercise of that very capacity. There is, however, an additional qualification in this case.

Given that an agent's background intrinsic desires might themselves lead him to interfere with his capacity to be instrumentally rational, or believe for reasons – think again about what the desire to have pleasurable experiences might lead an otherwise instrumentally rational agent to do – the required desires would have to be conditional in form. Fully rational agents would have to desire that they do not interfere with their exercise of their capacity to be instrumentally rational *on condition that*, by their exercising that very capacity, they do not form effective extrinsic desires to interfere with the exercise of their capacities to be instrumentally rational or believe for reasons. And note that this desire would also seem to be an instance of a more general desire that extends to all rational agents. Fully rational agents would have to desire that they do not interfere with any rational agent's exercise of their capacity to be instrumentally rational on condition that, by their exercising that capacity, those rational agents do not form effective extrinsic desires to interfere with the exercise of any other rational agent's, or their own, capacities to be instrumentally rational or believe for reasons (from here-on I will omit this qualification).

The desires described so far seem all to be instances of a pair of perfectly general desires whose content can be stated in the following terms: each fully rational agent desires not to interfere with any rational agent's exercise of his rational capacities, and they also desire that they do what they can to help agents have rational capacities to exercise. These desires, in turn, seem to be intrinsic, not extrinsic, because they don't depend on any belief that the things desired have some further feature that is desired. A fully rational agent simply has to desire these things themselves in order to function properly as a rational agent. The upshot is thus

that, even though Hume was right that all reasons are truth-supporting considerations, he was wrong that it follows from this that no intrinsic desires are required by reason. Agents are required by reason to have certain intrinsic desires because, absent their possession, they could not robustly possess and fully exercise their rational capacities. In particular, could not robustly possess and fully exercise a sensitivity to truth-supporting considerations in the formation of their beliefs.

Given that correct deliberation is a matter of the possession and exercise of rational capacities, this in turn has a crucial bearing on what agents would desire if they were to deliberate correctly. What they would desire that they do if they were to deliberate correctly turns out to be fixed not by what they would desire, whatever intrinsic desires they might happen to have, but by what they would desire, whatever intrinsic desires they might happen to have, if they in addition had the two intrinsic desires that are required by reason. Correct deliberation is thus *deliberation from*, inter alia, this pair of intrinsic desires.

3. Moral Obligations, Reasons for Action, and Agent-relative Values

As I said at the very beginning, I am sympathetic to the idea that any plausible moral theory will tell us that we have a moral obligation to produce states of affairs with more rather than less value. However, as I see things, the most plausible such theories will tell us that the values in question are one and all *non-welfarist* and *agent-relative*. It should by now be clear why this is so, but in case it isn't, let me make it explicit.

Any moral theory, to be in the least plausible, will have to tell us why moral obligations entail reasons for action. Given that an agent has a reason for action just in case he would desire that he so acts if he were to deliberate correctly, and given Hume's strictures about the rational status of intrinsic desires, this gives rise to the puzzle addressed in the last section, the solution to which is to acknowledge, as against Hume, that any agent, if he is robustly to possess and fully exercise rational capacities, must have the pair of intrinsic desires described: he must desire not to interfere with any rational agent's exercise of his rational capacities, and he must also desire that he does what he can to help agents have rational capacities to exercise. In virtue of the fact that every agent's fully rational

counterpart has these desires, every agent has the same reasons for action, and these reasons for action, I hereby conjecture, are reasons to do what agents are morally obliged to do. Agents are morally obliged not to interfere with any rational agent's exercise of his rational capacities, and they are also morally obliged to do what they can to make sure that agents have rational capacities to exercise.

Consider again the husband who has an obligation to treat his wife better than he does, but whose intrinsic desires are thoroughly nasty. The nastiness of his intrinsic desires provides no challenge at all to the idea that he has the two moral obligations just described. For in order to deliberate correctly, the imagined husband would have to robustly possess and fully exercise the capacities to believe for reasons and be instrumentally rational, and in order to robustly possess and fully exercise these capacities, he would have to intrinsically desire that he does not interfere with any rational agent's exercise of his rational capacities, and he would have to intrinsically desire to do what he can to make sure that agents have rational capacities to exercise. But if the husband deliberates from these desires – that is, if he forms extrinsic desires in the light of this pair of intrinsic desires – then he would desire that he does not interfere with his wife's exercise of her rational capacities, and he would also desire to do what he can to make sure that his wife possesses rational capacities. In other words, he would desire that he treats her much better than he does. All that the nastiness of his intrinsic desires would do is make him want to take steps to prevent them from ever being effective.

Are the two moral obligations just described grounded in values? It seems to me that they most certainly are. For according to the dispositional theory of value that I have argued for in earlier work, all that the desirability of some state of affairs, relative to some agent, consists in is that state of affairs' being the object of a desire that that agent would have if he were fully rational (Smith 1994). But since, as we have just seen, each fully rational agent would desire two things – that he does not interfere with any agent's exercise of his rational capacities and that he does what he can to make sure that agents possesses rational capacities – and given that each agent's desiring these two things is what explains why he has corresponding reasons for action and moral obligations, it follows that his moral obligations are grounded in values.

Are these values agent-relative or agent-neutral? The values in question are plainly agent-relative. They are agent-relative

because there is no way to characterize what is desirable without mentioning the agents themselves (Smith 2003). What each fully rational agent must desire, after all, is that *he himself* does not interfere with any rational agent's exercise of his capacities, not that every rational agent does so. And what each fully rational agent must also desire is that *he himself* does what he can to make sure that agents possesses rational capacities, not that every rational agent does what he can, and not that every agent possesses rational capacities whether or not anyone has to do anything at all to make sure that they possess them. What makes these states of affairs desirable, relative to each agent, is their being states of affairs in which the agent himself does not interfere with any rational agent's exercise of his capacities, or states of affairs in which the agent himself does what he can to make sure that agents possesses rational capacities. The values are thus agent-relative.

Are these agent-relative values welfarist or non-welfarist? The agent-relative values that ground the two moral obligations described above are plainly non-welfarist. For what turns out to be desirable is not pleasure as such, or the absence of pain, or anything else that constitutes an agent's welfare, but rather that an agent does what he can to make sure that agents possesses rational capacities, and that he does not interfere with any agent's exercise of his rational capacities. There will, of course, be a good deal of overlap between states of affairs in which (say) an agent does not interfere with any agent's exercise of his rational capacities, and those states of affairs in which there is an absence of pain, as causing pain is a very effective way to interfere with another agent's exercise of his rational capacities. Someone who has to focus all of his attention on dealing with pain typically lacks the psychic resources required to exercise his rational capacities. But the overlap is not perfect. If there are pains that have no effect on an agent's exercise of his rational capacities, then we have so far been given no reason to believe that these pains are undesirable, and if there are ways in which an agent's exercise of his rational capacities can be interfered with, but he suffers no loss of pleasure or welfare, then these acts of interference are undesirable even so. The values in question are thus plainly non-welfarist.

We are now in a position to see not just why it isn't true, but also why it is frankly implausible to suppose, that any plausible moral theory would include a principle of beneficence. A principle of beneficence purports to tell rational agents what they are morally obliged to do and hence what they have a reason to do. But given

that an agent has a reason to act in a certain way only if he would desire that he acts in that way if he were to deliberate correctly, it follows that if any plausible moral theory had to include a principle of beneficence, grounded in the agent-neutral value of welfare, then any agent who robustly possesses and fully exercises his capacity to believe for reasons and be instrumentally rational would have to desire that the world contains more welfare rather than less. But what is the connection supposed to be between such a free-floating concern for welfare and the robust possession and full exercise of rational capacities? The answer is that there is no connection at all. The ideas are orthogonal to each other. The same cannot be said, however, of the desires not to interfere with any agent's exercise of his rational capacities and to make sure that agents have rational capacities to exercise. As I have tried to argue, the connection between these desires and an agent's robust possession and full exercise of rational capacities is, more or less, transparent.

4. Conclusion

According to the abstract line of argument developed here, agents have two moral obligations. They are morally obliged not to interfere with any agent's exercise of his rational capacities and they are also morally obliged to do what they can to make sure that agents have rational capacities to exercise. Though these claims doubtless require much more in the way of defence than I have given them here, I hope I have said enough to make them sound at least plausible. If so, then it will have to be agreed that at least one plausible moral theory, the theory according to which agents have the two moral obligations just described, need not include a principle of beneficence, and, more generally, that such a theory could eschew both agent-neutral values and the moral obligations to which they would give rise. For though this theory does ground moral obligations in values, the values in question are all agent-relative and non-welfarist.

References

Arneson, Richard J. (2010). 'Good, Period', *Analysis* (70) pp. 731–44.
Broome, John (1991). *Weighing Goods* (Oxford: Blackwell).
Darwall, Stephen (1983). *Impartial Reason* (Ithaca: Cornell University Press).

—— (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge: Harvard University Press).
Dreier, James (1993). 'Structures of Normative Theories', *The Monist* (76) pp. 22–40.
Hume, David (1740). *A Treatise of Human Nature* (Oxford: Clarendon Press, 1740/1968).
Scanlon, Thomas M. (1998). *What We Owe To Each Other* (Cambridge: Harvard University Press).
Sen, Amartya (1982). 'Rights and Agency', *Philosophy and Public Affairs* (11) pp. 3–39.
—— (1988). 'Evaluator Relativity and Consequential Evaluation', in *Philosophy and Public Affairs* (12) pp. 113–132.
Smith, Michael (1994). *The Moral Problem* (Oxford: Blackwell Publishers).
—— (2003). 'Neutral and Relative Value after Moore', *Ethics,* Centenary Symposium on G.E.Moore's *Principia Ethica* (113) pp. 576–98.
—— (2004). 'Instrumental Desires, Instrumental Rationality', *Proceedings of the Aristotelian Society Supplementary Volume* (78) pp. 93–109.
—— (2009). 'Two Kinds of Consequentialism', *Philosophical Issues* (19), pp. 257–72.
—— (2010). 'Beyond the Error Theory', in *A World Without Values: Essays on John Mackie's Moral Error Theory* edited by Richard Joyce and Simon Kirchin (New York: Springer) pp. 119–39.
Williams, Bernard (1981). 'Internal and External Reasons', reprinted in his *Moral Luck* (Cambridge: Cambridge University Press).