

Don't Look Back in Anger: Retaining Moral Responsibility Without Free Will

Many feel that our practices of moral responsibility (holding others accountable, blaming and praising them and so on) would be fatally undermined if we did not have free will. I do not agree with this sceptical conclusion, and I will take a compatibilist approach to moral responsibility, characterised by Elinor Mason as: “[arguing] ...that although our actions are indeed ultimately caused by events outside of us, we can still make sense of moral responsibility.” (Mason, 2005, p343). I will develop a new concept of moral responsibility which I will call ‘reformatory moral responsibility’. I will contrast this with the ‘reactive attitudes’ proposed by PF Strawson (Strawson, 2003), and will conclude by showing how my ‘reformatory moral responsibility’ has greater coherence and explanatory power than the view of PF Strawson. For reasons of space, I will assume the truth of determinism, and the absence of free will, though I acknowledge that both of these propositions are a matter of considerable debate.

In holding someone morally responsible, and blaming them, it seems we have two linked thoughts. Firstly, we feel that something regrettable or bad has happened. Secondly, we assign responsibility for that event to a particular agent – a wrongdoer who has done wrong. As Smart puts it, blame involves “a grading plus an ascription of responsibility” (Smart, 2003, p70). And in Smart’s view, such non-judgemental ‘grading’ is justifiable, even when we have the belief that our actions are not fully under our own control. I will return to Smart’s point on grading later in this essay. In a recent paper, David Shoemaker has argued that any effective theory of moral responsibility must “incorporate and explain three distinct conceptions of responsibility – attributability, answerability, and accountability.” (Shoemaker, 2011). I will not go into the detail of Shoemaker’s distinction. Suffice to say that my account accommodates the strongest reading of Shoemaker’s definition of moral responsibility (i.e. all three conditions), even when agents are fully determined.

Now, to clear up some metaphysical confusion around the notion of ‘could have done otherwise’. When we blame others, we know their regrettable action has already happened – it is in the past. Following Dennett, we could wonder whether it makes any sense to ask ‘could they have done otherwise?’ (Dennett, 1984). Surely, contends Dennett, we don’t mean that if the wrongdoer were returned to exactly the *same* situation, equipped with exactly the *same* knowledge and motivations, they would then perform a *different* action to the one they actually did perform? Even if determinism were false, it is clear that we would expect a reliable agent to perform consistently. What

I mean by this is that if it were possible to return an agent to the exact circumstances (and mental state) leading up to their action, we would expect them to repeat the same action. If they did not, their behaviour would seem worryingly random or arbitrary. Therefore, 'could have done otherwise' is incoherent if it is interpreted in terms of the *same agent*, returned to the *same state*, performing a *different* action.

We must look for a more coherent and logically sound interpretation of the phrase 'could have done otherwise'. Interestingly, the only explanation Dennett explores is that asking 'could they have done otherwise?' is simply a practice of checking that no 'local fatalism' is in play – that the wrongdoer was unconstrained in the normal sense (Dennett, 1984 p554 and Dennett, 1984a Chapter 6). I agree with Dennett's view, namely that assigning responsibility for an act need not commit us to any metaphysical position about free will. As Dennett goes on to show with his 'deterministic robot' example, it can make perfect sense to describe a fully deterministic agent as 'faulty', and to examine ways to prevent such faults in future. Surprisingly, Dennett does not extend this analogy fully to interpersonal practices. Rather, when he considers interpersonal practices of blame, he falls back to the pragmatic "*holding people responsible is the best game in town*" (Dennett, 1984a, p162). I feel Dennett here misses a stronger and more meaningful approach to 'could have done otherwise'.

The stronger claim which Dennett could have made is to say – notwithstanding determinism – that holding people responsible is the *only* game in town. The basis for this claim is that as rational agents, we are responsive to reasons. And the condemnation or approval of others are reasons, just as much as our neurochemical makeup is a reason. To expand on Dennett, perhaps, when expressing the 'could have done otherwise' sentiment, what we really mean is that if our wrongdoer were returned to the scene of their crime with a *current* understanding of the consequences of their action (new knowledge, in effect) they would then refrain from their bad act. We might say something like 'now you know what hurt you caused with your remark, you should regret that you said it.' Therefore, a better way of understanding phrases such as 'could have done otherwise' is to view them as ways of identifying new reasons for the wrongdoer to avoid repeating their behaviour, or of giving them a new appreciation of old reasons – as much a case of education as condemnation. Understanding an ascription of moral responsibility in this way allows us to retain it, even when we acknowledge that everyone's actions are determined by factors beyond their complete control.

Therefore, I propose a new sense of moral responsibility, which I call 'reformatory moral responsibility'. Implicit in reformatory moral responsibility is the idea that an agent who has done wrong could not have done otherwise, based on their knowledge and motivations at the time. But we acknowledge the importance of working with the agent to ensure they are reformed for the future. Holding people morally responsible in this way has a rehabilitative function. Blame and sanctions can be thought of as human-produced features of our environment, hopefully altering our psychological makeup so that we are not the same person who once committed a crime. So even if an agent had no choice about how they acted in the sense of radical free will, we can still attribute the action to them – they are the proximate, if not the ultimate cause of their action. A reformatory approach to moral responsibility involves us in rationally engaging with the agent as the nearest relevant cause which we can alter or improve for future situations.

The notion of interpersonal condemnation as shaping the environment of an agent is touched on by J.J.C. Smart in his article 'Free Will, Praise and Blame' (Smart, 2003), where he gives the example of the justifiable punishment of a lazy schoolboy, in contrast to the unjustifiable punishment of his stupid classmate. Smart acknowledges that few would allow the schoolmaster to punish a 'stupid' boy, i.e. one genuinely incapable of doing his homework. However, says Smart:

On the other hand, the lazy boy can be influenced in such ways. Whether he does his homework or not is perhaps solely the outcome of environment, but one part of the environment is the threatening schoolmaster.

Threats and promises, punishments and rewards, the ascription of responsibility and the nonascription of responsibility, have therefore a clear pragmatic justification which is quite consistent with a wholehearted belief in metaphysical determinism. Indeed it implies a belief that our actions are very largely determined: if everything everyone did depended only on pure chance (i.e. if it depended on nothing) then threats and punishments would be quite ineffective.
– Smart, 2003, p68

Therefore, one can conceive of the schoolteacher's punishment as an attempt to reform the schoolboy, giving him new motivation to be more industrious in future. Smart later contrasts moralistic 'judging' – a negative practice – with the more favourable notion of 'grading'. In Smart's view, 'grading' is a justifiable form of evaluation, which shows more understanding of the agent's motivations. Although Smart does not spell out fully the reformatory possibilities offered by the combination of

'grading' and modified motivations, his example is a strong support for my view. Note also that Smart's account readily accommodates 'a wholehearted belief in metaphysical determinism' (Smart, 2003, p68).

I will now deal with a possible objection to my reformative view of moral responsibility. In 'Freedom and Resentment' (Strawson, 2003), PF Strawson contrasts various types of interpersonal attitudes. He favours the 'participant reactive attitudes' – i.e., ways in which we interact with other rational humans: holding them to account for their actions, praising and blaming them and so on. In cases where a person is compelled to behave badly (perhaps because of insanity or lack of moral development), Strawson suggests that we take a more distant 'objective attitude' towards him, suspending our normal practices of moral judgement. Strawson contends that if our moral practices were unduly influenced by determinism, all our moral judgements would have to fall under the 'objective attitude'. And, contends Strawson, this would undermine key parts of our common human experience. Therefore, one who accepted Strawson's account of objective and participative reactive attitudes might claim that my reformative characterisation of moral responsibility shares the flaws of the Strawsonian 'objective attitude'. Isn't there a danger that my reformative moral responsibility pathologises other agents, and treats them as mere objects to be, in Strawson's phrase "...handled, cured or trained..." (PF Strawson, 2003, p79)?

This objection highlights a legitimate worry, but it is ultimately unjustified. In my view, understanding other agents as constrained by determinism does not diminish their humanity. On the contrary, by treating their acts as inevitable given the agent's makeup, we can arrive at a truer understanding of what reasons they would need in order to behave differently in future. In understanding the binding and inevitable nature of the causes of their action, we are better able to find persuasive and motivating reasons for the agent to behave differently. (Here we are practising Smartian 'grading' in contrast to Strawsonian 'resentment'). With an appreciation of an agent's makeup, we avoid such dogmatic accusations as 'you should have known better!' – in my view, such expressions are incoherent. Given the act the agent *did* perform, it is nonsensical to say they knew (and were motivated to behave) better.

Another objection which could be levelled at my reformative account of blame is that claiming that a wrongdoer is merely badly informed is simply too weak, because it does not adequately account for our sense of anger when we have been wronged. And it does seem misguided to categorise a murderer, for example, as simply mistaken in their moral knowledge. It is true that reformative moral responsibility, in taking more

account of the reasons and causes which lead a wrongdoer to do something wrong might mean that we respond to wrongdoers with less anger. It is not clear that this is a drawback for my view – condemnation based on understanding rather than resentment seems progressive to me. The ‘missing anger’ objection could be put in terms of demanding an account of the role of anger in our practices of holding others accountable. Limitations of space mean that I cannot explore the role of anger in our moral practices here. Such an account is not needed for my main point, however, since one can easily imagine anger as supplying our initial motivation to intervene and remonstrate with agents who we feel have done wrong, while the more rational understanding of reformatory moral responsibility acts as a check to our sense of outrage, and ensures we intervene in an appropriate way. The criminal justice system, for example, puts much store in convicting and sentencing people in a dispassionate way, rather than acting out of indignation or a desire for vengeance.

One might find my account of reformatory moral responsibility convincing but still object to the threat posed by determinism to our sense of agency. In a determined universe, our objector might say, we are reduced to mere ‘happenings’ – and this is a threat to our sense of ourselves as agents. For reasons of space, I cannot respond to this kind of worry at length, but would draw such objectors’ attention to Nagel’s article ‘Freedom’ (Nagel, 2003), in which he gives a convincing treatment of this kind of problem. On Nagel’s account, regardless of the truth or falsity of determinism, this problem of ‘vanishing agency’ is a consequence of our point of view. Our own internal viewpoint is ‘agential’, yet thinking about the world at a highly reductive ‘external’ level has the effect of removing any sense of agency from the picture – things just seem to happen. One solution is to say that both pictures are right. A microphysical, causally determined picture is accurate in one (metaphysical) sense. Taken too far, though, we would be in danger of tracing every cause back to the origin of the universe. This might be true, but it is deeply trivial! Therefore, it makes sense to isolate a particular agent and hold them responsible for an action – we can reason with an agent, persuading them to behave differently next time. As noted above, we can hold someone responsible *qua* proximate cause, if not the ultimate cause of their action. This is a much more effective practice than attempting to alter their microphysical state, all the way back to the origin of the universe. As Dennett points out, since engineers can ‘average over the micro-details’ in analysing the performance of a deterministic robot, it is plausible for us to ignore much of the microphysical detail when interacting with each other (Dennett, 1984a, pp139–142).

Thus, treating people as agents need not require them to be self-causers or defy the laws of physics. And because humans have capacity to respond to psychological factors – reasons – and the ability to remember and reflect on previous cases of their own action, the practice of holding each other responsible remains effective and justified.

Conclusion

Building on Dennett’s critique of the coherence of ‘could have done otherwise’, I have developed an account of reformatory moral responsibility. I have drawn on Smart to show how moral interventions and sanctions remain coherent even in a deterministic universe. In holding people responsible and punishing them, we are literally *reforming* them – we aim to alter their motivational set so that if placed in similar circumstances in the future, they will behave differently. I have shown, *pace* Strawson, that such an ‘objective’ attitude need not alienate us from our interpersonal relations. Indeed, acknowledging the inevitability of actions which have been performed is likely to lead to a more effective practice of holding others accountable.

Bibliography

- Dennett, Daniel C. 1984: “I Could Not Have Done Otherwise –So What?” *The Journal of Philosophy* 81 (10) (October 1)
- 1984a, *Elbow Room: The Varieties of Free Will Worth Wanting*, Oxford: Clarendon Press
- Mason, Elinor 2005: ‘Recent work: Moral responsibility’. *Philosophical Books* 46 (4):343-353.
- Nagel, Thomas: 2003: ‘Freedom’ in Watson, Gary (Ed.) *Free Will*: Oxford, Oxford University Press
- Shoemaker, David: ‘Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility.’ *Ethics* Vol. 121, No. 3 (April 2011), pp. 602-632
- Smart, J.J.C.: 2003: ‘Free Will, Praise and Blame’ in Watson, Gary (Ed.) *Free Will*: Oxford, Oxford University Press
- Strawson, Peter: 2003: ‘Freedom and Resentment’ in Watson, Gary (Ed.) *Free Will*: Oxford, Oxford University Press

