

Free Will VI: Strawson & The Reactive Stance

Frankfurt argued that moral responsibility doesn't require the ability to do otherwise; so we can hold people responsible without answering the hard metaphysical question of whether they could have done otherwise. Strawson provides a very different way of trying to disconnect the two issues.

Strawson argues that there are two different sorts of attitudes that we can take towards a person: reactive attitudes and objective attitudes. As an example of the first, consider resentment. Suppose someone hurts you: pushes you over for instance. Then you will be apt to feel, not just angry, but resentful towards them. And this ties into a network of attitudes that goes much further than simply holding someone responsible: you might expect them to be sorry; you might forgive them if they do; your affection for them might be lost if they do not; and so on. However, if it turns out that they pushed you by accident these feelings will be diminished; perhaps they will go altogether; it depends on the nature of the case. If they had been rushing to try and help you, and had fallen into you they probably will; indeed they might even be replaced by positive reactive attitudes like gratitude. (Unless they have a history of clumsy over-solicitous interference.) If, on the other hand, they accidentally pushed you because they were recklessly trying to get past to see someone else, your feelings of resentment might be somewhat reduced, but they are unlikely to go completely. We require certain standards of good will and concern from those around us (how much depends on how close we are to them); and we feel resentment when we think these are not met. (Note: Strawson is rather unclear on what is required: normally he speaks of goodwill, but also of regard, affection and esteem.)

The objective attitude, in contrast, requires stepping back from such involvement. We move to it when we truly think that the harm was caused accidentally; here we view the *action* objectively. In these cases we go on treating the *person* reactively. But sometimes we no longer take the reactive stance to a person in just about everything they do. Consider, for instance, someone suffering from serious mental illnesses like schizophrenia, or severe depression. Here we adopt the objective stance to the *person*. We no longer feel resentment to them; we rather think of them as people needing to be treated or managed. Why do we suspend the reactive stance when dealing with the very young or the mentally ill? Again because we do not think that their actions indicate a lack of good will towards us.

How should we understand these different sorts of attitude? Strawson's first point is that they are attitudes on the part of *the person who is reacting*, not different facts about the person they are reacting to. This is shown by the fact that we sometimes take both attitudes towards the same person for the same acts. Thus: suppose you are involved in a torrid love affair. Most of the time you are passionately in the grip of reactive attitudes to your beloved: affection, resentment as you think they scorn you, affection again as they ask your forgiveness, jealousy as you think they are unfaithful, and so on. Sometimes though it all gets too much. To escape from the unbearable strain you fall back into the objective stance. You remind yourself of the terrible childhood they had; you think of the earlier affairs they had that left them so wounded; you explain what they are doing as a result of the insecurities that these experiences have inflicted upon them. You can't maintain this suspension of the reactive stance for very long (or at least you can't suspend it and maintain the relationship with them; once it's over you might think of their actions, and indeed your own, almost entirely objectively); but the fact that you can suspend it at all shows that taking the stance is something that isn't just dictated by them; it's something that comes, at least partly, from you.

Strawson now asks: why should we think that acknowledging the truth of determinism will make us give up the reactive stance across the board? He thinks that it is absurd to think that it would. In the first place, we are just not able to give it up; it is too deeply ingrained. It would be to give up on our humanity. Secondly, even if we ask whether the acknowledging the truth of determinism *should rationally* make us give it up (whether or not we actually would), the answer is: no. For the decision as to whether we should rationally take the reactive stance is a practical one; whereas the truth of determinism is a theoretical question. Practical questions are answered by considering how we would benefit from the different answers to them. And there is no question that our lives would be hugely impoverished if we were to give up on the reactive attitudes. (The difference between practical and theoretical questions is clearly related to the difference between *is*-statements and *ought*-statements. Roughly theoretical question as questions about how things are; and practical questions are questions about what we ought to do; and there are good reasons for thinking that we can never derive an *ought*-statement from an *is*-statement; the contention sometimes known as *Hume's Law*. Note too that one can think there is a fairly clear divide between *is*-statements and *ought*-statements without thinking that there is a clear divide between fact and value.)

All of this has addressed the compatibility of determinism and our practice of ascribing moral responsibility. But what about the compatibilism of determinism with the claim that people are free? (What we might call *freedom compatibilism*, to distinguish it from *responsibility compatibilism*.) There are two ways to go here. One is to accept that the claim that people are free is a theoretical claim, but one which is detached from the practical question of how we should treat people. So even if freedom compatibilism is false, this won't have any terrible consequences for our practice of ascribing blame. (But what should we make of the phenomenology of freedom?)

There is another way to understand Strawson. This is to understand him as saying that the claim that a certain person is free is not really a descriptive theoretical claim at all. Rather it is simply something we say when we are prepared to take the reactive stance towards them. (Compare *expressivism* or *emotivism* in ethics: expressivists say that the statement that a given thing is good isn't really a descriptive statement at all. It's just something that we say that expresses our attitude to it; it's like saying 'Hurrah'.) Understood this way, Strawson comes out as a kind of sophisticated freedom compatibilist: the claim that determinism is true is quite compatible with the claim that people have free will, since the latter is just a way of indicating that we are prepared to take the reactive stance towards them. It is not a descriptive statement either. (But again, what now becomes of the phenomenology of freedom?) Yet another approach would liken Strawson to a subjectivist: to claim that someone is free is to claim that one will be taking the subjective stance towards them. So this is a descriptive statement, but it is about oneself.

Independently of this, we might wonder how successful Strawson's approach is on its own terms. And the big worry here is whether or not we can isolate the theoretical from the practical in the way he suggests. One aspect of this is the following: don't we think that there is a real question of whether we are right in thinking that we are justified in holding someone responsible? We don't just want to say: 'This is what we do'; we want to say 'This is what we are justified in doing'.

One way of thinking about this is to compare it with other practices. Thus suppose someone made a parallel defense of religion. Suppose it could be shown that there was a natural human tendency to believe in a god, and that human beings flourished when they did. Would that justify belief in god? Wouldn't we want to say: whether or not it is unavoidable,

or it benefits us, we want to know whether the belief is *justified*. We want to know whether there is a god.

WATSON: RESPONSIBILITY AND THE LIMITS OF EVIL

Recall that Strawson gave two sorts of circumstance under which we would not normally take the reactive stance:

Type One Conditions: the agent was pushed, didn't realize the harm that their action would cause, was acting under terrible duress etc. The condition doesn't generalize to other actions.

Type Two Conditions: the agent was scarcely an agent: they were a child, mentally ill, a psychopath etc.; or there was a similar but temporary excuse: they were under great strain, depressed etc. The condition does generalize to other actions.

The relevance of Type One Conditions is easily explained on Strawson's account: since, in exhibiting reactive attitudes, what we are reacting to is a lack of good will or regard on the part of the agent, we should not form such attitudes where the agent's action does not manifest such a lack. But what of Type Two Conditions: a child, or someone under stress, or a psychopath can manifest ill will towards me. So why should I suspend the reactive stance?

The traditional theorist has something to say: being in these situations removes one's moral responsibility. But Strawson can't simply say this, for to have moral responsibility just is to be someone to whom the reactive stance is taken. So that would not be an explanation of why we do not take the reactive stance in Type Two Conditions; rather it follows from the fact that we do not. We need to know what the rationale is for not taking that stance.

Watson offers one suggestion: to take the reactive stance is partly a communicative act: it involves a form of address. Perhaps the idea is that people under Type Two Conditions are not people with whom we can meaningfully communicate.

Watson takes the case of Robert Harris, who carried out some horrific murders, but who had an equally horrific upbringing. When we hear what he did, we are horrified; but when we learn about his childhood, this makes a difference to our response. It is not that we fail to take the reactive stance at all; but it is somehow qualified. ('We are less inclined to use certain *reactive* epithets such as "scumbag"')

Can the 'address' approach explain this? It is far from clear that it can. For whether or not we can meaningfully communicate with someone is a function of how they are now. How they came to be this way doesn't seem relevant.

Incompatibilists will contend that this provides support for their account. For isn't this just confirmation of their hypothesis that we no longer hold someone responsible once we see them as determined by their background? But Watson thinks that this isn't right either since: (i) we don't give up on the reactive stance altogether; and (ii) it doesn't seem that we need to think of Harris as compelled by his childhood into his subsequent life: 'The thought is not "It had to be!" but, again, "No wonder!"' (Is this right: couldn't the incompatibilists' real point be put in terms of causation rather than determinism?)

VARDAS: REVISIONISM

One general question about Strawson—indeed a general question about all accounts of moral responsibility or of free will—concerns how much we are trying to accommodate the folk intuitions. Compare science or mathematics. We might have all kinds of beliefs about the world (that heavier things fall faster than lighter things, for instance, or that there are twice as many natural numbers as there are even natural numbers) that turn out to be wrong. But we still think that we are talking about the same things. Similarly, perhaps we just have some false metaphysical beliefs about moral responsibility—that it requires the ability to do otherwise or suchlike. If so, perhaps we can revise our account to eliminate the falsehoods. A difficult question here is why we're not just changing the subject: giving up on moral responsibility and replacing it with moral responsibility*. In the case of science it is the world that provides the continuity. What is it for moral responsibility (or indeed for maths)?

One interesting parallel here: views on race. It's not clear that there is any scientific foundation for an ontology of human races, yet many people feel that their identity is tightly bound up with their race, and so many theorists in the area want to maintain the notion. Presumably the criteria we use to apply the notion will be partly driven by normative considerations.

Can we do the same with moral responsibility? And with free will? Revisions want to do something like that: to hang on to what we want and jettison the rest, falling somewhere between compatibilism and hard determinism.