

On Free Will *Galen Strawson*

Are we free agents? Can we be morally responsible for what we do? Some philosophers answer Yes and Yes (we are fully free, and wholly morally responsible for what we do). Others answer Yes and No (certainly we are free agents - but we cannot be said to be *ultimately* responsible for what we do). A third group answers No and No (we are not free agents at all; *a fortiori* we cannot be morally responsible). A strange minority answers No and Yes (we can be morally responsible for what we do, even though we are not free agents!). This last view is rare, but it has a kind of existentialist panache, and appears to be embraced by Wintergreen in Joseph Heller's novel *Closing Time* (1994), as well as by some Protestants.

Who's right? Suppose tomorrow's a holiday, and that you're wondering what to do. You can climb a mountain or read Lao Tzu. You can restring your mandolin or go to the zoo. At the moment you're reading about free will. You're free to go on reading or stop now. You have started on this sentence, but you don't have to.....finish it. Right now, as so often in life, you have a number of options. Nothing forces your hand. So surely you're entirely free to choose what to do, and responsible for what you do?

This is what the 'Compatibilists' think. They answer Yes and Yes. Their name derives from their claim that free will is entirely compatible with determinism (the view that everything that happens in the universe is necessitated by what has already gone before, in such a way that nothing can happen otherwise than it does). Free will, they think, is just a matter of not being constrained or compelled in certain ways that have nothing to do with whether determinism is true or false. 'Consider yourself at this moment', they say. 'No one's holding a gun to your head. You're not being threatened or manhandled. You're not drugged, or in chains, or subject to a psychological compulsion like kleptomania, or a post-hypnotic command. So you're wholly free. This is what being a free agent is. It's wholly irrelevant that your character is determined, if indeed it is.

'And although things like guns and chains, threats to the life of your children, psychological obsessions, and so on, are standardly counted as constraints that can limit freedom and responsibility, there is another and more fundamental sense in which you are fully free in any situation in which you can choose or act in any way at all - in any situation in which you are not panicked, or literally forced to do what you do. Consider pilots of hijacked aeroplanes. They usually stay calm. They choose to comply with the hijackers' demands. They act responsibly, as we naturally say. They are able to do other than they do, but they choose not to. They do what they most want to do, all things considered, in the circumstances in which they find themselves - and all circumstances limit one's options in some way. Some circumstances limit one's options much more drastically than others, but it doesn't follow that one isn't free to choose in those circumstances. Only literal compulsion, panic, or uncontrollable impulse really removes one's freedom to choose, and to (try to) do what one most wants to do, given one's character or personality. Even when one's finger is being forced down on the button, one can still act freely in resisting the pressure, in cursing one's oppressor, and in many other ways.'

So most of us are wholly free to choose and act throughout our waking lives, according to the Compatibilists. We're free to choose between the options we perceive to be open to us.

One has options even when one is in chains, or falling through space. Even if one is completely paralysed, one is still free in so far as one is free to choose to think about one thing rather than another. There is, as Sartre observed, a sense in which we're condemned to freedom: we're not free not to be free. One may well not be able to do everything one wants - one may want to fly unassisted, vapourize every gun in the United States by an act of thought, or house all those who sleep on the streets of Calcutta by the end of the month - but few have supposed that free will is a matter of being able to do everything one wants. It's a possible view, but according to the Compatibilists free will is simply a matter of being unconstrained in such a way that one has genuine options and opportunities for action, and is able to choose between them according to what one wants or thinks best. It just doesn't matter if one's character, personality, preferences, and general motivational set are entirely determined by things for which one is in no way responsible - by one's genetic inheritance, upbringing, historical situation, chance encounters, and so on.

Even dogs count as free agents, on this view. So Compatibilists have to explain what distinguishes us from dogs - since we don't think that dogs are free in the way we are. Some say it's our capacity to act for reasons that we explicitly take to be moral reasons. Many say it's our capacity for explicitly self-conscious thought. Not because self-consciousness liberates anyone from determinism; if determinism is true, one is determined to have whatever self-conscious thoughts one has, whatever their complexity.

The idea is that self-consciousness makes it possible for one to be explicitly aware of oneself as facing choices and engaging in processes of reasoning about what to do, and thereby constitutes one as a radically free agent in a way unavailable to any unself-conscious agent. They think one's self-conscious deliberative presence in the situation of choice simply trumps the fact - if it is a fact - that one is in the final analysis wholly constituted as the sort of person one is by factors for which one isn't in any way ultimately responsible. The Compatibilists, then, say Yes and Yes, and those who want to say this are well advised to follow them, for determinism is unfalsifiable, and may be true. (In the end, contemporary physics gives us no more reason to suppose that determinism is false than to suppose that it is true.)

Many, however, think that Compatibilism doesn't even touch the real problem of free will. For what is it to define freedom in such a way that it is compatible with determinism? It's to define it in such a way that an agent can be a free agent even if all its actions throughout its life are determined to happen as they do by events that have taken place before it is born, so that there is a clear sense in which it could not at any point in its life have done otherwise than it did. And this doesn't look like genuine free will or moral responsibility. How can one possibly be truly or ultimately morally responsible for what one does, if everything one does is ultimately a deterministic outcome of events for whose occurrence one is in no way responsible? This is the *Incompatibilists'* view.

The Incompatibilists divide into two groups: there are the *Libertarians*, on the one hand, and the *No-Freedom Theorists* or *Pessimists*, on the other. The Libertarians are up-beat. They say Yes and Yes, like the Compatibilists, but think the Compatibilists' account of freedom can be improved. They hold (1) that we do have free will, (2) that free will is not compatible with determinism, and (3) that determinism is therefore false. But then they face an extremely difficult task: they have to show how indeterminism (the falsity of determinism) can help with free will, and in particular with moral responsibility.

The Pessimists don't think this can be shown. They agree that free will is not compatible with determinism, but deny that indeterminism can help. They think that free will, of the sort that is necessary for genuine moral responsibility, is provably impossible. They say No and No.

They begin by granting what everyone must. They grant that there's a clear, important, compatibilist sense in which we can be free agents (we can be free, when unconstrained, to choose and to do what we want or think best, given how we are). But they insist that this isn't enough: it doesn't give us what we want in the way of free will. Nor does it give us what we believe we have. But (they continue) it is not as if the Compatibilists have missed something. The truth is that nothing can give us what we think we want, and ordinarily think we have. We cannot be morally responsible, in the absolute, buck-stopping way in which we often unreflectively think we are. We cannot have 'strong' free will of the kind that we would need to have, in order to be morally responsible in this way.

One way of setting out the Pessimists' argument is as follows: (1) When you act, you do what you do, in the situation in which you find yourself, because of the way you are. But then (2) To be truly or ultimately morally responsible for what you do, you must be truly or ultimately responsible for the way you are, at least in certain crucial mental respects. (Obviously you don't have to be responsible for your height, age, sex, and so on.)

But (3) You can't be ultimately responsible for the way you are in any respect at all, so you can't be ultimately responsible for what you do.

For (4) To be ultimately responsible for the way you are, you must have somehow intentionally brought it about that you are the way you are. And the problem is then this. Suppose (5) You have somehow intentionally brought it about that you are the way you now are, in certain mental respects: suppose you have brought it about that you have a certain mental nature Z, in such a way that you can be said to be ultimately responsible for Z.

For this to be true (6) You must already have had a certain mental nature Y, in the light of which you brought it about that you now have Z. (If you didn't already have a mental nature then you didn't have any intentions or preferences, and can't be responsible for the way you now are, even if you have changed.) But then (7) For it to be true that you are ultimately responsible for how you now are, you must be ultimately responsible for having had that nature, Y, in the light of which you brought it about that you now have Z.

So (8) You must have brought it about that you had Y.

But then (9) You must have existed already with a prior nature, X, in the light of which you brought it about that you had Y, in the light of which you brought it about that you now have Z.

And so on. Here, one is setting off on a potentially infinite regress. In order for one to be truly or ultimately responsible for how one is in such a way that one can be truly responsible for what one does, something impossible has to be true: there has to be, and cannot be, a starting point in the series of acts of bringing it about that one has a certain nature; a starting point that constitutes an act of ultimate self-origination. There's a more concise way of putting the point: in order to be ultimately responsible, one would have to be *causa sui* - the ultimate cause or origin of oneself, or at least of some crucial part of one's mental nature. But nothing can be ultimately *causa sui* in any respect at all. Even if the property of being *causa sui* is allowed to belong unintelligibly to God, it cannot plausibly be supposed to be possessed by ordinary finite human beings. 'The *causa sui* is the best self-contradiction that has been conceived so far', as Nietzsche remarked in *Beyond Good and Evil* in 1886:

it is a sort of rape and perversion of logic. But the extravagant pride of man has managed to entangle itself profoundly and frightfully with just this nonsense. The desire for 'freedom of the will' in the superlative metaphysical sense, which still holds sway, unfortunately, in the minds of the half-educated; the desire to bear the entire and ultimate responsibility for one's actions oneself, and to absolve God, the world, ancestors, chance, and society involves nothing less than to be precisely this *causa sui* and, with more than Baron Muenchhausen's audacity, to pull oneself up into existence by the hair, out of the swamps of nothingness. . . .

In fact, nearly all of those who believe in strong free will do so without any conscious thought that it requires ultimate self-origination. But self-origination is the only thing that could actually ground the kind of strong free will that is regularly believed in.

The Pessimists' argument may seem contrived, but essentially the same argument can be given in a more natural form as follows. (A) One is the way one is, initially, as a result of heredity and early experience. (B) These are clearly things for which one cannot be held to be in any way responsible (this might not be true if there were reincarnation, but this would just shift the problem backwards). (C) One cannot at any later stage of one's life hope to accede to ultimate responsibility for the way one is by trying to change the way one already is as a result of heredity and experience. For one may well try to change oneself, but (D) both the particular way in which one is moved to try to change oneself, and the degree of one's success in one's attempt at change, will be determined by how one already is as a result of heredity and experience. And (E) any further changes that one can bring about only after one has brought about certain initial changes will in turn be determined, via the initial changes, by heredity and previous experience. (F) This may not be the whole story, for it may be that some changes in the way one is are traceable to the influence of indeterministic or random factors. But (G) it is absurd to suppose that indeterministic or random factors, for which one is *ex hypothesi* in no way responsible, can in themselves contribute to one's being truly or ultimately responsible for how one is. The claim, then, is not that

people cannot change the way they are. They can, in certain respects (which tend to be exaggerated by North Americans and underestimated, perhaps, by members of other cultures). The claim is only that people cannot be supposed to change themselves in such a way as to be or become ultimately responsible for the way they are, and hence for their actions. One can put the point by saying that in the final analysis the way you are is, in every last detail, a matter of luck - good or bad.

Philosophers will ask what exactly this 'ultimate' responsibility is supposed to be. They will suggest that it doesn't really make sense, and try to move from there to the claim that it can't really be what we have in mind when we talk about moral responsibility. It is very clear to most people, however, and one dramatic way to characterize it is by reference to the story of heaven and hell: it is responsibility of such a kind that, if we have it, it makes sense to propose that it could be just to punish some of us with torment in hell and reward others with bliss in heaven. It makes sense because what we do is absolutely up to us. The words 'makes sense' are stressed, because one doesn't have to believe in the story of heaven and hell in order to understand the notion of ultimate responsibility that it is used to illustrate. Nor does one have to believe in it in order to believe in ultimate responsibility (many atheists have done so).

The story is useful, because it illustrates the kind of absolute or ultimate responsibility that many have supposed - and do suppose - themselves to have. (Another way to characterize it is to say that it exists if punishment and reward can be fair without having any pragmatic - or indeed aesthetic - justification.) But one doesn't have to appeal to it when describing the sorts of everyday situation that are primarily influential in giving rise to our belief in ultimate responsibility. Suppose you set off for a shop on the evening of a national holiday, intending to buy a cake with your last £10 note. Everything is closing down. There is one cake left; it costs £10. On the steps of the shop, someone is shaking an Oxfam tin. You stop, and it seems completely clear to you that it is entirely up to you what you do next: you are truly, radically free to choose, in such a way that you will be ultimately responsible for whatever you do choose. You can put the money in the tin, or go in and buy the cake, or just walk away. You are not only completely free to choose. You are not free not to choose. Standing there, you may believe determinism is true: you may believe that in five minutes time you will be able to look back on the situation you are now in and say, of what you will by then have done, 'It was determined that I should do that'. But even if you do wholeheartedly believe this, it does not seem to touch your current sense of the absoluteness of your freedom and moral responsibility.

One diagnosis of this phenomenon is that one can't really believe that determinism is true in such situations, and also can't help thinking that its falsity might make freedom possible. But the feeling of ultimate responsibility seems to remain inescapable even if this is not so. Suppose one fully accepts the Pessimists' argument that no one can be *causa sui*, and that one has to be *causa sui* (in certain crucial mental respects) in order to be ultimately responsible for one's actions. This does not seem to have any impact on one's sense of one's radical freedom and responsibility, as one stands there, wondering what to do. One's radical responsibility seems to stem simply from the fact that one is fully conscious of one's situation, and knows that one can choose, and believes that one action is morally better than the other. This seems to be immediately enough to confer full and ultimate responsibility. And yet it cannot

really do so, according to the Pessimists. For whatever one actually does, one will do what one does because of the way one is, and the way one is is something for which one neither is nor can be responsible, however self-consciously aware of one's situation one is.

The Pessimists' argument is hard to stomach (even Hitler is let off the hook), and one challenge to it runs as follows. 'Look, the reason why one can be ultimately responsible for what one does is that one's self is, in some crucial sense, independent of one's general mental nature (character or motivational structure). Suppose one faces a difficult choice between A, doing one's moral duty, and B, following one's desires. You Pessimists describe this situation as follows. Given one's mental nature, you say, one responds in a certain way. One is swayed by reasons for and against both A and B. One tends towards A or B, and in the end one does one or the other, given one's mental nature, which is something for which one cannot be ultimately responsible. But this description of yours forgets the self - it forgets what one might call 'the agent-self'. As an agent-self, one is in some way independent of one's mental nature. One's mental nature inclines one to do one thing rather than another, but it does not thereby necessitate one to do one thing rather than the other (to use Leibniz's terms).

As an agent-self, one incorporates a power of free decision that is independent of all the particularities of one's mental nature in such a way that one can, after all, count as ultimately morally responsible in one's decisions and actions, even though one is not ultimately responsible for any aspect of one's mental nature.'

The Pessimists are unimpressed: 'Even if one grants the validity of this conception of the agent-self for the sake of argument', they say, 'it cannot help. For if the agent-self decides in the light of the agent's mental nature but is not determined by the agent's mental nature, the following question immediately arises: Why does the dear old agent-self decide as it does? The general answer is clear. Whatever it decides, it decides as it does because of the overall way it is, and this necessary truth returns us to where we started: somehow, the agent-self is going to have to get to be responsible for being the way it is, in order for its decisions to be a source of ultimate responsibility. But this is impossible: nothing can be *causa sui* in the required way. Whatever the nature of the agent-self, it is ultimately a matter of luck. Maybe the agent-self decides as it does partly or wholly because of the presence of indeterministic occurrences in the decision process. Maybe, maybe not. It makes no difference, for indeterministic occurrences can never contribute to ultimate moral responsibility.' Some think they can avoid this by asserting that free will and moral responsibility are just a matter of being governed by reason - or by Reason with a dignifying capital 'R'. But being governed by Reason can't be the source of ultimate responsibility. It can't be a property that makes punishment ultimately just or fair for those who possess it, and unfair for those who don't. For to be morally responsible, on this view, is simply to possess one sort of motivational set among others. But if you do possess this motivational set, then you are simply lucky - if it is indeed a good thing - while those who lack it are unlucky.

This will be denied. It will be said, truly, that some people struggle to become more morally responsible, and make an enormous effort. Their moral responsibility is then not a matter of luck; it's their own hard-won achievement. The Pessimists' reply is

immediate. 'Suppose you are someone who struggles to be morally responsible, and make an enormous effort. Well, that too is a matter of luck. You are lucky to be someone who has a character of a sort that disposes you to make that sort of effort. Someone who lacks a character of that sort is merely unlucky.' In the end, luck swallows everything. This is one (admittedly contentious) way of putting the point that there can be no ultimate responsibility, given the natural, strong conception of responsibility that was characterized by reference to the story of heaven and hell. Relative to that conception, no punishment or reward is ever ultimately just or fair, however natural or useful or otherwise humanly appropriate it may be or seem.

Galen Strawson
University of Reading

(Note: A version of this paper first appeared in *The Times Literary Supplement*.)