

FREE WILL

Galen Strawson

Abstract

'Free will' is the conventional name of a topic that is best discussed without reference to the will. It is a topic in metaphysics and ethics as much as in the philosophy of mind. Its central questions are 'What is it to act (or choose) freely?', and 'What is it to be morally responsible for one's actions (or choices)?' These two questions are closely connected, for it seems clear that freedom of action is a necessary condition of moral responsibility, even if it is not sufficient.

Philosophers give very different answers to these questions. Consequently they give very different answers to two more specific questions, which are questions about ourselves: (1) Are we free agents? and (2) Can we be morally responsible for what we do? Answers to (1) and (2) range from 'Yes, Yes', to 'No, No'-via 'Yes, No' and various degrees of 'Perhaps', 'Possibly', and 'In a sense'. (The fourth pair of outright answers, 'No, Yes', is rare, but it has a kind of existentialist panache, and appears to be embraced by Wintergreen in Joseph Heller's novel *Closing Time*, as well as by some Protestants).

Prominent among the 'Yes, Yes' sayers are the compatibilists. They have this name because they hold that free will is compatible with determinism. Briefly, determinism is the view that the history of the universe is fixed: everything that happens is necessitated by what has already gone before, in such a way that nothing can happen otherwise than it does. According to compatibilists, freedom is compatible with determinism because freedom is essentially just a matter of not being constrained or hindered in certain ways when one acts or chooses. Suppose one is a normal adult human being in normal circumstances. Then one is able to act and choose freely. No one is holding a gun to one's head. One is not being threatened or manhandled. One is not drugged, or in chains, or subject to a psychological compulsion like kleptomania, or a post-hypnotic command. One is therefore wholly free to choose and act even if one's whole physical and psychological makeup is entirely determined by things for which one is in no way ultimately responsible-starting with one's genetic inheritance and early upbringing.

Compatibilism has many sophisticated variants, but this is its core, and to state it is to see what motivates its opponents, the incompatibilists. The incompatibilists hold that freedom is not compatible with determinism. They point out that if determinism is true, then every one of one's actions was determined to happen as it did before one was born. They hold that one can't be held to be truly free and finally morally responsible for one's actions in this case. Compatibilism is a 'wretched subterfuge..., a petty word-jugglery', as Kant put it. It entirely fails to satisfy our natural convictions about the nature of moral responsibility.

The incompatibilists have a good point, and may be divided into two groups. First, there are the libertarians, who wish to answer 'Yes, Yes' to questions (1) and (2). Libertarians hold that we are indeed free and fully morally responsible agents, and that determinism must therefore be false. Their great difficulty is to explain why the falsity of determinism is any better than determinism, when it comes to establishing our free agency and moral responsibility. For suppose that not every event is determined, and that some events occur randomly, or as a matter of chance. How can this help with free will? How can our claim to moral responsibility be improved by the supposition that it is partly a matter of chance or random outcome that we and our actions are as they are? This is a very difficult question for libertarians.

The second group of incompatibilists are less sanguine. They answer 'No, No' to questions (1) and (2). They agree with the libertarians that determinism rules out genuine moral responsibility, but argue that the falsity of determinism can't help. Accordingly, they conclude that we are not genuinely free agents or genuinely morally responsible, whether determinism is true or false. One of their arguments can be summarized as follows. When one acts, one acts in the way that one does because of the way one is. So to be truly morally responsible for one's actions, one would have to be truly responsible for the way one is: one would have to be *causa sui*, or the cause of oneself, at least in certain crucial mental respects. But nothing can be *causa sui* - nothing can be the ultimate cause of itself in any respect. So nothing can be truly morally responsible.

Suitably developed, this argument against moral responsibility seems very strong. But in many human societies belief in ultimate moral responsibility continues unabated. In many human beings, the experience of choice gives rise to a conviction of absolute responsibility that is untouched by philosophical arguments that put it in question.

This conviction is the deep and inexhaustible source of the free will problem: there are powerful arguments that seem to show that we cannot be morally responsible in the ultimate way that we suppose. But these arguments keep coming up against equally powerful psychological and cultural reasons why we continue to believe that we are ultimately morally responsible.

1. Compatibilism

Do we have free will? It depends what you mean by the word 'free'. More than 200 senses of the word have been distinguished; the history of the discussion of free will is rich and remarkable. David Hume called the problem of free will "the most contentious question of metaphysics, the most contentious science" (Enquiry p. 95).

Here it will be enough to focus on two main senses of the word 'free'. The first has the consequence that the answer to the question 'Do we have free will?' is 'Yes'. The second has the consequence that the answer is 'No'. The first is compatibilist; that is, it is a sense of the word 'free' according to which free will is compatible with determinism, even though determinism is the view that the history of the universe is fixed in such a way that nothing can happen otherwise than it does, because everything that happens is necessitated by what has already gone before.

Suppose tomorrow is a national holiday. You are considering what to do. You can climb a mountain or read Lao Tzu. You can mend your bicycle or go to the zoo. At this moment you are reading this article. You are free to go on reading or stop now. You have started on this sentence, but you don't have to finish it.

In this situation, as so often in life, you have a number of options. Nothing forces your hand. It seems natural to say that you are entirely free to choose what to do. And, given that nothing hinders you, it seems natural to say that you act entirely freely when you actually do (or try to do) what you have decided to do.

Compatibilists claim that this is the right thing to say. They believe that to have free will, to be a free agent, to be free in choice and action, is simply to be free from constraints of certain sorts. Freedom is a matter of not being physically or psychologically forced or compelled to do what one does.

This raises the question of what counts as a constraint or compulsion. In one sense, compatibilist freedom can be limited by imprisonment, which is likely to prevent one from doing what one wants to do. It can be limited by a gun at one's head, or a threat to the life of one's children, or a psychological obsession. All these things are standardly counted as constraints that can limit freedom.

In another and more fundamental sense, however, compatibilist freedom is something one continues to possess undiminished so long as one can choose or act in any way at all. One continues to possess it in any situation in which one is not panicked, or literally compelled to do what one does in such a way that it is not clear that one can still be said to choose or act at all (as when one presses a button, because one's finger is actually forced down on the button).

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.

It is true that 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, and in many other ways.

So most of us are free to choose throughout our waking lives, according to the compatibilist conception of freedom. We are free to choose between the options that we perceive to be open to us. (Sometimes we would rather not face options, but are unable to avoid awareness of the fact that we do face them.) 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. Sartre observed that there is a sense in which we are 'condemned' to freedom, 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 or free agency is a matter of being able to do everything one wants. That is one possible view of what it is to be free; but according to the compatibilists, free will is simply a matter of having genuine options and opportunities for action, and being able to choose between them according to what one wants or thinks best.

Compatibilists grant that one's character, personality, preferences, and general motivational set may be entirely determined by things for which one is in no way responsible. These things may be determined, for example, by one's genetic inheritance, upbringing, historical situation, chance encounters, and so on. But one does not have to be in control of any of these things in order to have compatibilist freedom, because compatibilist freedom is just a matter of being able to choose and act in the way one prefers or thinks best given how one is. As its name declares, it is compatible with determinism. It is compatible with determinism even though it follows from determinism that every aspect of your character, and everything you will ever do, was already inevitable before you were born.

It may be said that dogs and other animals can be free agents, according to this basic account of compatibilism. Compatibilists may reply that dogs can indeed be free agents. And yet we do not think that dogs can be free or morally responsible in the way we can be. So compatibilists need to say what the relevant difference is between dogs and ourselves.

Many suppose that it is our capacity for self-conscious thought that makes the crucial difference, because it makes it possible for us to be explicitly aware of ourselves as facing choices and engaging in processes of reasoning about what to do. This is not because being self-conscious can somehow liberate one from the facts of determinism: if determinism is true, one is determined to have whatever self-conscious thoughts one has, whatever their complexity. Nevertheless, many are inclined to think that a creature's explicit self-conscious awareness of itself as chooser and agent can constitute it as a free agent in a fundamental way that is unavailable to any unself-conscious agent.

Compatibilists can agree with this. They can acknowledge and incorporate the view that self-conscious awareness of oneself as facing choices can give rise to a kind of freedom that is unavailable to unself-conscious agents. They may add that human beings are sharply marked off from dogs by their capacity to act for reasons that they explicitly take to be moral reasons. In general, compatibilism has many variants. According to H. Frankfurt's version, for example, one has free will if one wants to be moved to action by the motives that do in fact move one to action. On this view, freedom is a matter of having a personality that is harmonious in a certain way. Freedom in this sense is clearly compatible with determinism.

Compatibilism has been refined in many ways, but this gives an idea of its basis. 'What more could free agency possibly be?', compatibilists like to ask (backed by Hobbes, Locke, and Hume, among others). And this is a very powerful question.

2. Incompatibilism

Those who want to secure the conclusion that we are free agents do well to adopt a compatibilist theory of freedom, for determinism is unfalsifiable, and may be true. (Contemporary physics gives us no more reason to suppose that determinism is false than to suppose that it is true.) Many, however, think that the compatibilist account of things does not even touch the real problem of free will. They believe that all compatibilist theories of freedom are patently inadequate.

What is it, they say, to define freedom in such a way that it is compatible with determinism? It is 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. This, they say, is certainly not free will. More importantly, it is not a sufficient basis for true moral responsibility. One cannot possibly be truly or ultimately morally responsible for what one does if everything one does is ultimately a deterministic outcome of events that took place before one was born; or (more generally) a deterministic outcome of events for whose occurrence one is in no way ultimately responsible.

These anti-compatibilists or incompatibilists divide into two groups: the libertarians and the no-freedom theorists or pessimists about free will and moral responsibility. The libertarians think that the compatibilist account of freedom can be improved on. They hold (1) that we do have free will, (2) that free will is not compatible with deter-

minism, and (3) that determinism is therefore false. But 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 or no-freedom theorists do not think that this can be shown. They agree with the libertarians that the compatibilist account of free will is inadequate, but they don't think it can be improved on. They agree that free will is not compatible with determinism, but deny that indeterminism can help to make us (or anyone else) free. They believe that free will, of the sort that is necessary for genuine moral responsibility, is provably impossible.

The pessimists about free will begin by granting what everyone must. They grant that there is a clear and 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 go on to insist that this compatibilist sense of freedom 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. And it is not as if the compatibilists have missed something. The truth is that nothing can give us what we (think we) want, or what we ordinarily think we have. All attempts to furnish a stronger notion of free will fail. 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.

It is the worry about moral responsibility that is the fundamental motor of the free will debate. If no one had this worry, it is doubtful whether the problem of free will would be a famous philosophical problem. The rest of this discussion will therefore be organized around the question of moral responsibility.

First, though, it is worth remarking that the worry about free will does not have to be expressed as a worry about the grounds of moral responsibility. Two points are worth making. The first is that a commitment to belief in free will may be integral to feelings that are extremely important to us independently of the issue of moral responsibility: feelings of gratitude, for example, and perhaps of love. The second is that one's belief in strong free will may be driven simply by the conviction that one is or can be radically self-determining in one's actions in a way that is incompatible with determinism; and this conviction about radical self-determination need not involve giving much-or any-thought to the issue of moral responsibility. It seems that a creature could conceive of itself as radically self-determining without having any conception of moral right or wrong at all-and so without being any sort of moral agent.

3. Pessimism

One way of setting out the no-freedom theorists' 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.

It seems to follow that

(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 the way you are in all respects. You don't have to be responsible for your height, age, sex, and so on. But it does seem that you have to be responsible for the way you are at least in certain mental respects. After all, it is your overall mental make up that leads you to do what you do when you act.)

But

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

(4) To be ultimately responsible for the way you are, you would have to have intentionally brought it about that you are the way you are, in a way that is impossible.

The impossibility is shown as follows. Suppose that

(5) You have somehow intentionally brought it about that you are the way you now are, in certain mental respects: suppose that you have intentionally brought it about that you have a certain mental nature *N*, and that you have brought this about in such a way that you can now be said to be ultimately responsible for having nature *N*. (The

limiting case of this would be the case in which you had simply endorsed your existing mental nature N from a position of power to change it.)

For this to be true

(6) You must already have had a certain mental nature $N - 1$, in the light of which you intentionally brought it about that you now have nature N . (If you didn't already have a certain mental nature, then you can't have had any intentions or preferences, and even if you did change in some way, you can't be held to be responsible for the way you now are.)

But then

(7) For it to be true that you and you alone are truly responsible for how you now are, you must be truly responsible for having had the nature $N - 1$ in the light of which you intentionally brought it about that you now have nature N .

So

(8) You must have intentionally brought it about that you had that nature $N - 1$. But in that case, you must have existed already with a prior nature, $N - 2$, in the light of which you intentionally brought it about that you had the nature $N - 1$ in the light of which you intentionally brought it about that you now have nature N .

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 morally 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 is a more concise way of putting the point: in order to be truly morally responsible for what one does, it seems that one would have to be the ultimate cause or origin of oneself, or at least of some crucial part of one's mental nature. One would have to be *causa sui*, in the old terminology. But nothing can be truly or 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 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 Münchhausen's audacity, to pull oneself up into existence by the hair, out of the swamps of nothingness...' (Beyond Good and Evil, §21)

In fact, nearly all of those who believe in strong free will do so without any conscious thought that it requires ultimate self-origination. It remains true that such self-origination is the only thing that could actually ground the kind of strong free will that is regularly believed in. And it does seem that one way in which the belief in strong free will manifests itself is in the very vague and (necessarily) unexamined belief that many have that they are somehow or other radically responsible for their general mental nature, or at least for certain crucial aspects of it.

The pessimists' argument may seem contrived, but essentially the same argument can be given in a more natural form as follows.

(1) It is undeniable that one is the way one is, initially, as a result of heredity and early experience.

(2) It is undeniable that these are things for which one cannot be held to be in any way responsible (this might not be true if there were reincarnation, but reincarnation would just shift the problem backwards).

(3) One cannot at any later stage of one's life hope to accede to true or ultimate responsibility for the way one is by trying to change the way one already is as a result of one's heredity and previous experience.

For one may well try to change oneself, but

(4) 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 previous experience.

And

(5) 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.

(6) 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

(7) it is foolish 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 many other cultures). The claim is only that people cannot be supposed to change themselves in such a way as to be or become truly or ultimately responsible for the way they are, and hence for their actions. One can put the point by saying that the way you are is, ultimately, in every last detail, a matter of luck-good or bad.

4. Moral responsibility

Two main questions are raised by the pessimists' arguments. First, is it really true that one needs to be self-creating or *causa sui* in some way, in order to be truly or ultimately responsible for what one does, as step (2) of the pessimists' argument asserts? This question will be delayed until §6, because a more basic question arises: What notion of responsibility is being appealed to in this argument? What exactly is this 'ultimate' responsibility that we are held to believe in, in spite of Nietzsche's scorn? And if we do believe in it, what makes us believe in it?

One dramatic way to characterize the notion of ultimate responsibility is by reference to the story of heaven and hell: 'ultimate' moral responsibility 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 certainly does not 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 the story of heaven and hell in order to believe in ultimate responsibility (many atheists have believed in it). One doesn't have to have heard of it.

The story is useful because it illustrates the kind of absolute or ultimate responsibility that many have supposed and do suppose-themselves to have. It is particularly vivid when one is specifically concerned with moral responsibility, and with questions of desert (punishment and reward), but it serves equally well to illustrate the sense of radical freedom and responsibility that may be had by a self-conscious agent that has no concept of morality. And one does not have to refer to the story of heaven and hell in order to describe the sorts of everyday situation that are perhaps 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 ten pound note. Everything is closing down. There is one cake left; it costs ten pounds. 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. That is, it seems clear to you that 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 that 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 believe this, it does not seem to undermine your current sense of the absoluteness of your freedom, and of your moral responsibility for your choice.

One diagnosis of this phenomenon is that one can't really believe that determinism is true, in such situations of choice, and can't help thinking that the falsity of determinism might make freedom possible. But the feeling of ultimate responsibility seems to remain inescapable even if one doesn't think this, and even if one has been con-

vinced by the entirely general argument against ultimate responsibility given in §3. Suppose one accepts that no one can be in any way *causa sui*, and that one would have 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 example of the cake may be artificial, but similar situations of choice occur regularly in human life. They are the experiential rock on which the belief in ultimate responsibility is founded. The belief often takes the form of belief in specifically moral, desert-implying responsibility. But an agent could have a sense of ultimate responsibility without possessing any conception of morality, as noted, and there is an interesting intermediate case: an agent could have an irrepressible experience of ultimate responsibility, and believe in objective moral right and wrong, while still denying the coherence of the notion of desert.

5 Metaphysics and moral psychology

We now have the main elements of the problem of free will. It is natural to start with the compatibilist position; but this has only to be stated to trigger the objection that compatibilism cannot possibly satisfy our intuitions about moral responsibility. According to this objection, an incompatibilist notion of free will is essential in order to make sense of the idea that we are genuinely morally responsible. But this view, too, has only to be stated to trigger the pessimists' objection that indeterministic occurrences cannot possibly contribute to moral responsibility: one can hardly be supposed to be more truly morally responsible for one's choices and actions or character if indeterministic occurrences have played a part in their causation than if they have not played such a part. Indeterminism gives rise to unpredictability, not responsibility. It cannot help in any way at all.

The pessimists therefore conclude that strong free will is not possible, and that ultimate responsibility is not possible either. So no punishment or reward is ever truly just or fair, when it comes to moral matters.

This conclusion may prompt a further question: What exactly is this 'ultimate' responsibility that we are supposed to believe in? One answer refers to the story of heaven and hell, which serves to illustrate the kind of responsibility that is shown to be impossible by the pessimists' argument, and which many people do undoubtedly believe themselves to have, however fuzzily they think about the matter. A less colourful answer has the same import, although it needs more thought: 'ultimate' responsibility exists if and only if punishment and reward can be fair without having any pragmatic justification.

Now the argument may cycle back to compatibilism. Pointing out that that 'ultimate' moral responsibility is obviously impossible, compatibilists may claim that we should rest content with the compatibilist account of things—since it is the best we can do. But this claim reactivates the incompatibilist objection, and the cycle continues.

There is an alternative strategy at this point: quit the traditional metaphysical circle for the domain of moral psychology. The principal positions in the traditional metaphysical debate are clear. No radically new option is likely to emerge after millennia of debate. The interesting questions that remain are primarily psychological: Why do we believe we have strong free will and ultimate responsibility of the kind that can be characterized by reference to the story of heaven and hell? What is it like to live with this belief? What are its varieties? How might we be changed by dwelling intensely on the view that ultimate responsibility is impossible? And so on.

A full answer to these questions is beyond the scope of this article, but one fundamental cause of our belief in ultimate responsibility has been mentioned. It lies in the experience of choice that we have as self-conscious agents who are able to be fully conscious of what they are doing when they deliberate about what to do and make choices. (We choose between the Oxfam box and the cake; or make a difficult, morally neutral choice about which of two paintings to buy.) This raises an interesting question: Is it true that any self-conscious agent that faces choices and is fully aware of the fact that it does so must experience itself as having strong free will, or as being radically self-determining, simply in virtue of the fact that it is a self-conscious agent (and whether or not it has a conception of moral responsibility)? It seems that we cannot live or experience our choices as determined, even if determinism is true. But perhaps this is a human peculiarity, not an inescapable feature of any possible self-conscious agent. And perhaps it is not even universal among human beings.

Other causes of the belief in strong free will have been suggested. Hume stressed our experience of serious indecision, as above. Spinoza proposed that one of the causes is simply that we are not conscious of the determined nature of our desires. Kant held that our experience of moral obligation makes belief in strong free will inevitable. P. F. Strawson argued that the fundamental fact is that we are irresistibly committed to certain natural reactions to other people like gratitude and resentment. Various other suggestions have been made: those who think hard about free will are likely to become convinced that investigation of the complex moral psychology of the belief in freedom, and of the possible moral and psychological consequences of altering the belief, is the most fruitful area of research that remains. New generations, however, will continue to launch themselves onto the old metaphysical roundabout.

6 Challenges to pessimism

The preceding discussion attempts to illustrate the internal dynamic of the free will debate, and to explain why the debate is likely to continue for as long as human beings can think. The basic point is this: powerful logical or metaphysical reasons for supposing that we can't have strong free will keep coming up against equally powerful psychological reasons why we can't help believing that we do have it. The pessimists' or no-freedom theorists' conclusions may seem irresistible during philosophical discussion, but they are likely to lose their force, and seem obviously irrelevant to life, when one stops philosophizing.

Various challenges to the pessimists' argument have been proposed, some of which appear to be supported by the experience or 'phenomenology' of choice. One challenge grants that one cannot be ultimately responsible for one's mental nature - one's character, personality, or motivational structure-but denies that it follows that one can't be truly morally responsible for what one does (it therefore challenges step (2) of the argument set out in §3).

This challenge has at least two versions. One has already been noted: we are attracted by the idea that our capacity for fully explicit self-conscious deliberation, in a situation of choice, suffices by itself to constitute us as truly morally responsible agents in the strongest possible sense. The idea is that such full self-conscious awareness somehow renders irrelevant the fact that one neither is nor can be ultimately responsible for any aspect of one's mental nature. On this view, the mere fact of one's self-conscious presence in the situation of choice can confer true moral responsibility: it may be undeniable that one is, in the final analysis, wholly constituted as the sort of person one is by factors for which one cannot be in any way ultimately responsible; but the threat that this fact appears to pose to one's claim to true moral responsibility is simply obliterated by one's self-conscious awareness of one's situation.

The pessimists reply: This may correctly describe a strong source of belief in ultimate (moral) responsibility, but it is not an account of something that could constitute ultimate (moral) responsibility. When one acts after explicit self-conscious deliberation, one acts for certain reasons. But which reasons finally weigh with one is a matter of one's mental nature, which is something for which one cannot be in any way ultimately responsible. One can certainly be a morally responsible agent in the sense of being aware of distinctively moral considerations when one acts. But one cannot be morally responsible in such a way that one is ultimately deserving of punishment or reward for what one does.

The conviction that fully explicit self-conscious awareness of one's situation can be a sufficient foundation of strong free will is extremely powerful. The no-freedom theorists' argument seems to show that it is wrong, but it is a conviction that runs deeper than rational argument, and it survives untouched, in the everyday conduct of life, even after the validity of the no-freedom theorists' argument has been admitted.

Another version of the challenge runs as follows. The reason why one can be truly or ultimately (morally) responsible for what one does is that one's self is, in some crucial sense, independent of one's general mental nature (one's character, personality, motivational structure, and so on). Suppose one faces a difficult choice between A, doing one's duty, and B, following one's non-moral desires. The pessimists describe this situation as follows: Given one's mental nature, they 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.

Those who challenge this description say that it reckons without the self - without 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. (The distinction between inclining and necessitating derives from Leibniz.) 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 truly and 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 reply: Even if one grants the validity of this conception of the agent-self for the sake of argument, it cannot help to establish ultimate moral responsibility. According to the conception, 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 agent-self decide as it does? The general answer is clear. Whatever the agent-self decides, it decides as it does because of the overall way it is; and this necessary truth returns us to where we started. For once again, it seems that the agent-self must be responsible for being the way it is, in order to be a source of true or ultimate responsibility. But this is impossible, for the reasons given in §3: nothing can be *causa sui* in the required way. Whatever the nature of the agent-self, it is ultimately a matter of luck (or, for those who believe in God, a matter of grace). It may be proposed that the agent-self decides as it does partly or wholly because of the presence of indeterministic occurrences in the decision process. But it is clear that indeterministic occurrences can never be a source of true (moral) responsibility.

Some believe that free will and moral responsibility are above all a matter of being governed in one's choices and actions by reason-or by Reason with a capital 'R'. But possession of the property of being governed by Reason cannot be a ground of radical moral responsibility as ordinarily understood. It cannot be a property that makes punishment (for example) ultimately just or fair for those who possess it, and unfair for those who do not possess it. Why not? Because to be morally responsible, on this view, is simply to possess one sort of motivational set among others. It is to value or respond naturally to rational considerations - which are often thought to include moral considerations, by those who propound this view. It is to have a general motivational set that may be attractive, and that may be more socially beneficial than many others. But there is no escape from the fact that someone who does possess such a motivational set is simply lucky to possess it - if it is indeed a good thing-while someone who lacks it is unlucky.

This may be denied. It may be said 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 is 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. Kant is a famous example of a philosopher who was attracted by the idea that to display free will is to be governed by Reason in one's actions. But he became aware of the problem just described, and insisted, in a later work, that 'man himself must make or have made himself into whatever, in a moral sense, whether good or evil, he is to become. Either condition must be an effect of his free choice; for otherwise he could not be held responsible for it and could therefore be morally neither good nor evil'. Since he was committed to belief in ultimate moral responsibility, Kant held that such self-creation does indeed take place, and wrote accordingly of 'man's character, which he himself creates', and of 'knowledge [that one has] of oneself as a person who ... is his own originator'. Here he made the demand for self-creation that is natural for someone who believes in ultimate moral responsibility and who thinks through what is required for it.

In the end, luck swallows everything. This is one way of putting the point that there can be no ultimate responsibility, given the natural, strong conception of responsibility that was characterized at the beginning of §4. 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.

The facts are clear, and they have been known for a long time. When it comes to the metaphysics of free will, André Gide's remark is apt: 'Everything has been said before, but since nobody listens we have to keep going back and beginning all over again'. It seems that the only freedom that we can have is compatibilist freedom. If, since that is not enough for ultimate responsibility, we cannot have ultimate responsibility. The only alternative to this conclusion is to appeal to God and mystery - this in order to back up the claim that something that appears to be provably impossible is not only possible but actual.

The debate continues; some have thought that philosophy ought to move on. There is little reason to expect that it will do so, as each new generation arises bearing philosophers gripped by the conviction that they can have ultimate responsibility. Would it be a good thing if philosophy did move on, or if we became more clear-headed about the topic of free will than we are? It's hard to say.