

Chapter Two

The Impossibility of Ultimate Responsibility?

Galen Strawson

1. THE BASIC ARGUMENT

You set off for a shop on the evening of a national holiday, intending to buy a cake with your last five-pound note to supplement the preparations you've already made. There's one cake left in the shop and it costs five pounds; everything is closing down. On the steps of the shop someone is shaking a box, collecting money for Oxfam. You stop, and it seems clear to you that it is *entirely up to you* what you do next. It seems clear to you that you are truly, radically free to choose, in such a way that you will be ultimately morally responsible for whatever you do choose.

There is, however, an argument, which I will call the Basic Argument, which appears to show that we can never be truly or ultimately morally responsible for our actions. According to the Basic Argument, it makes no difference whether determinism is true or false.

The central idea can be quickly conveyed:

- a. Nothing can be *causa sui*—nothing can be the cause of itself.
- b. In order to be truly or ultimately morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects.
- c. Therefore no one can be truly or ultimately morally responsible.

We can expand the argument as follows:

1. Interested in free action, we're particularly interested in actions performed for a reason (as opposed to reflex actions or mindlessly habitual actions).
2. When one acts for a reason, what one does is a function of how one is, mentally speaking. (It's also a function of one's height, one's strength, one's place and time, and so on; but it's the mental factors that are crucial when moral responsibility is in question.)
3. So if one is to be truly responsible for how one acts, one must be truly responsible for how one is, mentally speaking—at least in certain respects.
4. But to be truly responsible for how one is, in any mental respect, one must have brought it about that one is the way one is, in that respect. And it's not merely that one must have caused oneself to be the way one is, in that respect. One must also have consciously and explicitly chosen to be the way one is, in that respect, and one must have succeeded in bringing it about that one is that way.
5. But one can't really be said to choose, in a conscious, reasoned, fashion, to be the way one is in any respect at all, unless one already exists, mentally speaking, already equipped with some principles of choice, "P1"—preferences, values, ideals—in the light of which one chooses how to be.
6. But then to be truly responsible, on account of having chosen to be the way one is, in certain mental respects, one must be truly responsible for one's having the principles of choice P1 in the light of which one chose how to be.
7. But for this to be so one must have chosen P1, in a reasoned, conscious, intentional fashion.
8. But for this to be so one must already have had some principles of choice P2, in the light of which one chose P1.
9. And so on. Here we are setting out on a regress that we cannot stop. True self-determination is impossible because it requires the actual completion of an infinite series of choices of principles of choice.
10. So true moral responsibility is impossible, because it requires true self-determination, as noted in (3).¹

This may seem contrived, but essentially the same argument can be given in a more natural form.

1. It's undeniable that one is the way one is, initially, as a result of heredity and early experience, and it's undeniable that these are things for which one can't be held to be in any way responsible (morally or otherwise).

2. One can't at any later stage of life hope to accede to true or ultimate moral responsibility for the way one is by trying to change the way one already is as a result of one's genetic inheritance and previous experience.

For

3. 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 to change, will be determined by how one already is as a result of one's genetic inheritance and previous experience.

And

4. 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 one's genetic inheritance and previous experience.
5. This may not be the whole story, and there may be changes in the way one is that can't be traced to one's genetic inheritance and experience but rather to the influence of indeterministic factors. It is, however, absurd to suppose that indeterministic factors, for which one is obviously not responsible, can contribute in any way to one's being truly morally responsible for how one is.

2. ULTIMATE MORAL RESPONSIBILITY

But what is this supposed "true" or "ultimate" moral responsibility? An old story may be helpful. As I understand it, it's responsibility of such a kind that, if we have it, then it *makes sense* to suppose that it could be just to punish some of us with (eternal) torment in hell and reward others with (eternal) bliss in heaven. The stress on the words "makes sense" is important, because one certainly doesn't have to believe in any version of the story of heaven and hell in order to understand, or indeed believe in, the kind of true or ultimate moral responsibility that I'm using the story to illustrate. A less colorful way to convey the point, perhaps, is to say that true or ultimate responsibility exists if punishment and reward can be fair without having any sort of pragmatic justification whatever.

One certainly doesn't have to refer to religious faith in order to describe the sorts of everyday situation that give rise to our belief in such responsibility. Choices like the one with which I began (the cake or the collection box) arise all the time, and constantly refresh our conviction about our responsibility. Even if one believes that determinism is true, in such a situation, and that one will in five minutes' time be able to look back and say that what one did was determined, this doesn't seem to undermine one's sense of the absolute-

ness and inescapability of one's freedom, and of one's moral responsibility for one's choice. Even if one accepts the validity of the Basic Argument, which concludes that one can't be in any way ultimately responsible for the way one is and decides, one's freedom and true moral responsibility seem, in the moment, as one stands there, obvious and absolute.

Large and small, morally significant or morally neutral, such situations of choice occur regularly in human life. I think they lie at the heart of the experience of freedom and moral responsibility. They're the fundamental source of our inability to give up belief in true or ultimate moral responsibility. We may wonder why human beings experience these situations of choice as they do; it's an interesting question whether any possible cognitively sophisticated, rational, self-conscious agent must experience situations of choice in this way.² But these situations of choice are the experiential rock on which the belief in ultimate moral responsibility is founded.

Most people who believe in ultimate moral responsibility take its existence for granted, and don't ever entertain the thought that one needs to be ultimately responsible for the way one is in order to be ultimately responsible for the way one *acts*. Some, however, reveal that they see its force. E. H. Carr states that "normal adult human beings are morally responsible for their own personality" (Carr 1961, 89). Sartre holds that "man is responsible for what he is" (Sartre 1946, 29) and seeks to give an account of how we "choose ourselves" (Sartre 1943, 440, 468, 503). In a later interview he judges his earlier assertions about freedom to be incautious—"When I read this, I said to myself: it's incredible, I actually believed that!"—but he still holds that "in the end one is always responsible for what is made of one" (Sartre 1970, 22). Kant puts it clearly when he claims 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. (Kant 1793, 40)

Since he is committed to belief in radical moral responsibility, Kant holds that such self-creation does indeed take place, and writes accordingly of "man's character, which he himself creates" (Kant 1788, 101), and of "knowledge of oneself as a person who . . . is his own originator" (Kant 1800, 213). John Patten, a former British Secretary of State for Education, claims that "it is . . . self-evident that as we grow up each individual chooses whether to be good or bad" (Patten 1992).³ Robert Kane, an eloquent recent defender of this view, writes as follows:

if . . . a choice issues from, and can be sufficiently explained by, an agent's character and motives (together with background conditions), then to be ultimately responsible for the choice, the agent must be at least in part responsible

by virtue of choices or actions voluntarily performed in the past for having the character and motives he or she now has. (Kane 2000, 317–18)

Christine Korsgaard agrees: “judgements of responsibility don’t really make sense unless people create themselves” (Korsgaard 2009, 20).

Most of us, as remarked, never actually follow this line of thought. It seems, though, that we do tend, in some vague and unexamined fashion, to think of ourselves as responsible for—answerable for—how we are. The point is somewhat delicate, for we don’t ordinarily suppose that we have gone through some sort of active process of self-determination at some past time. It seems nevertheless that we do unreflectively experience ourselves, in many respects, rather as we might experience ourselves if we did believe that we had engaged in some such activity of self-determination; and we may well also think of others in this way.

Sometimes a part of one’s character—a desire or tendency—may strike one as foreign or alien. But it can do this only against a background of character traits that aren’t experienced as foreign, but are rather “identified” with. (It’s only relative to such a background that a character trait can stand out as alien.) Some feel tormented by impulses that they experience as alien, but in many a sense of general identification with their character predominates, and this identification seems to carry within itself an implicit sense that one is generally speaking in control of, or at least answerable for, how one is (even, perhaps, for aspects of one’s character that one doesn’t like). So it seems that we find, semi-dormant in common thought, an implicit recognition of the idea that true or ultimate moral responsibility for one’s actions (for what one does) does somehow involve responsibility for how one is: it seems that ordinary thought is ready to move this way under pressure.

There are also many aspects of our ordinary sense of ourselves as morally responsible free agents that we don’t feel to be threatened in any way by the fact that we can’t be ultimately responsible for how we are. We readily accept that we are products of our heredity and environment without feeling that this poses any threat to our freedom and moral responsibility at the time of action. It’s natural to feel that if one is fully consciously aware of oneself as able to choose in a situation of choice, then this is already entirely sufficient for one’s radical freedom of choice—whatever else is or is not the case (see further the penultimate paragraph of this chapter). It seems, then, that our ordinary conception of moral responsibility may contain mutually inconsistent elements. If this is so, it is a profoundly important fact (it would explain a great deal about the character of the philosophical debate about free will). But these other elements in our ordinary notion of moral responsibility, important as they are, are not my present subject.⁴

3. RESTATEMENT OF THE BASIC ARGUMENT

I want now to restate the Basic Argument in very loose—as it were conversational—terms. New forms of words allow for new forms of objection, but they may be helpful nonetheless—or for that reason.

1. You do what you do, in any situation in which you find yourself, because of the way you are.

So

2. To be truly morally responsible for what you do you must be truly responsible for the way you are—at least in certain crucial mental respects.

Or:

1. When you act, what you do is a function of how you are (what you do won't count as an action at all unless it flows appropriately from your beliefs, preferences, and so on).

Hence

2. You have to get to have some responsibility for how you are in order to get to have some responsibility for what you intentionally do.

Once again I take the qualification about “certain mental respects” for granted. Obviously one isn't responsible for one's sex, basic body pattern, height, and so on. But if one weren't responsible for anything about oneself, how could one be responsible for what one did, given the truth of (1)? This is the fundamental question, and it seems clear that if one is going to be responsible for any aspect of oneself, it had better be some aspect of one's mental nature.

I take it that (1) is incontrovertible, and that it is (2) that must be resisted. For if (1) and (2) are conceded the case seems lost, because the full argument runs as follows:

1. You do what you do because of the way you are.⁵

So

2. To be truly morally responsible for what you do you must be truly responsible for the way you are—at least in certain crucial mental respects.

But

3. You can't be truly responsible for the way you are, so you can't be truly responsible for what you do.

Why can't you be truly responsible for the way you are? Because

4. To be truly responsible for the way you are, you must have intentionally brought it about that you are the way you are, and this is impossible.

Why is it impossible? Well, suppose it isn't. Suppose

5. You have somehow intentionally brought it about that you are the way you now are, and that you have brought this about in such a way that you can now be said to be truly responsible for being the way you are now.

For this to be true

6. You must already have had a certain nature N in the light of which you intentionally brought it about that you are as you now are.

But then

7. For it to be true that you are truly responsible for how you now are, you must be truly responsible for having had the nature N in the light of which you intentionally brought it about that you are the way you now are.

So

8. You must have intentionally brought it about that you had that nature N, in which case you must have existed already with a prior nature in the light of which you intentionally brought it about that you had the nature N in the light of which you intentionally brought it about that you are the way you now are . . .

Here one is setting off on the regress again. Nothing can be *causa sui* in the required way. Even if this attribute is allowed to belong (unintelligibly) to God, it can't 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 . . . (Nietzsche 1886, §21)

The rephrased argument is essentially exactly the same as before, although the first two steps are now more simply stated. Can the Basic Argument simply be dismissed? Is it really of no importance in the discussion of free will and moral responsibility, as some have claimed? (No and No.) Shouldn't any serious defense of free will and moral responsibility thoroughly acknowledge the respect in which the Basic Argument is valid before going on to try to give its own positive account of the nature of free will and moral responsibility? Doesn't the argument go to the heart of things if the heart of the free will debate is a concern about whether we can be truly morally responsible in the absolute way that we ordinarily suppose? (Yes and Yes.)

We are what we are, and we can't be thought to have made ourselves *in such a way* that we can be held to be free in our actions *in such a way* that we can be held to be morally responsible for our actions *in such a way* that any punishment or reward for our actions is ultimately just or fair. Punishments and rewards may seem deeply appropriate or intrinsically "fitting" to us; many of the various institutions of punishment and reward in human society appear to be practically indispensable in both their legal and non-legal forms. But if one takes the notion of justice that is central to our intellectual and cultural tradition seriously, then the consequence of the Basic Argument is that there is a fundamental sense in which no punishment or reward is ever just. It is exactly as just to punish or reward people for their actions as it is to punish or reward them for the (natural) color of their hair or the (natural) shape of their faces. The conclusion seems intolerable, but inescapable.

Darwin develops the point as follows in a notebook entry for 6 September 1838:

The general delusion about free will obvious. . . . One must view a [wicked] man like a sickly one—We cannot help loathing a diseased offensive object, so we view wickedness.—it would however be more proper to pity than to hate & be disgusted. with them. Yet it is right to punish criminals; but solely to *deter* others. . . . This view should teach one profound humility, one deserves no credit for anything. (yet one takes it for beauty and good temper), nor ought one to blame others.—This view will do no harm, because no one can be really *fully* convinced of its truth. except man who has thought very much, & he will know his happiness lays in doing good & being perfect, & therefore will not be tempted, from knowing every thing he does is independent of himself[,] to do harm.⁶

4. RESPONSE TO THE BASIC ARGUMENT

I've suggested that it is step (2) of the restated Basic Argument that must be rejected, and of course it can be rejected, because the phrases "truly responsible" and "truly morally responsible" can be defined in many ways. I'll sketch three sorts of response.

(I) The first response is *compatibilist*. Compatibilists say that one can be a free and morally responsible agent even if determinism is true. They claim that one can correctly be said to be truly responsible for what one does, when one acts, just so long as one is in control of one's action in the way that we take an ordinary person to be in ordinary circumstances: one isn't, for example, caused to do what one does by any of a certain set of constraints (kleptomaniac impulses, obsessional neuroses, desires that are experienced as alien, post-hypnotic commands, threats, instances of *force majeure*, and so on). Compatibilists don't impose any requirement that one should be truly responsible for how one is, so step (2) of the Basic Argument comes out as false, on their view. They think one can be fully morally responsible even if the way one is is entirely determined by factors entirely outside one's control. They simply reject the Basic Argument. They know that the kind of responsibility ruled out by the Basic Argument is impossible, and conclude that it can't be the kind of responsibility that is really in question in human life, because (they insist) we are indeed genuinely morally responsible agents. No theory that concludes otherwise can possibly be right, on their view.

(II) The second response is *libertarian*. *Incompatibilists* believe that freedom and moral responsibility are incompatible with determinism, and some incompatibilists are libertarians, who believe that we are free and morally responsible agents, and that determinism is therefore false. Robert Kane, for example, allows that we may act responsibly from a will already formed, but argues that the will must in this case be

“our own” free will by virtue of other past “self-forming” choices or other actions that were undetermined and by which we made ourselves into the kinds of persons we are . . . [T]hese undetermined self-forming actions (SFAs) occur at those difficult times of life when we are torn between competing visions of what we should do or become. (Kane 2000, 318–19)

They paradigmatically involve a conflict between moral duty and non-moral desire, and it is essential that they involve indeterminism, on Kane's view, for this “screens off complete determination by influences of the past” (Kane 2000, 319). He proposes that we are in such cases of “moral, prudential and practical struggle . . . truly ‘making ourselves’ in such a way that we are ultimately responsible for the outcome,” and that this “making of ourselves” means that “we can be ultimately responsible for our present motives and character by virtue of past choices which helped to form them and for which we were ultimately responsible” (Kane 1989, 252).

Kane, then, accepts step (2) of the Basic Argument, and challenges step (3) instead. He accepts that we have to “make ourselves,” and so be ultimately responsible for ourselves, in order to be morally responsible for what we do; and he thinks that this requires indeterminism. But the old, general objec-

tion to libertarianism recurs. How can indeterminism possibly help with moral responsibility? How can the occurrence of indeterministic or partly random events contribute to my being truly or ultimately morally responsible either for my actions or for my character? If my efforts of will shape my character in a positive way, and are in so doing essentially partly indeterministic in nature, while also being shaped (as Kane grants) by my already existing character, why am I not merely *lucky*?

This seems to be a very strong general objection to any libertarian account of free will. Suppose, in the light of this, that we put aside the Basic Argument for a moment, and take it as given that there is—that there must be—some respectable sense in which human beings are or can be genuinely morally responsible for their actions. If we then ask what sort of account of moral responsibility this will be, compatibilist or incompatibilist, I think we can safely reply that it will have to be compatibilist. This is because it seems so clear that nothing can ever be gained, in an attempt to defend moral responsibility, by assuming that determinism is false.

(III) The third response begins by accepting that one can't be held to be ultimately responsible for one's character or personality or motivational structure. It accepts that this is so whether determinism is true or false. It then directly challenges step (2) of the Basic Argument. It appeals to a certain picture of the *self* in order to argue that one can be truly free and morally responsible in spite of the fact that one can't be held to be ultimately responsible for one's character or personality or motivational structure.

It can be set out as follows. One is free and truly morally responsible because one's self is, in a crucial sense, independent of one's character or personality or motivational structure—one's CPM, for short. Suppose one is in a situation which one experiences as a difficult choice between A, doing one's duty, and B, following one's non-moral desires. Given one's CPM, one responds in a certain way. One's desires and beliefs develop and interact and constitute reasons in favor both of A and of B, and one's CPM makes one tend towards either A or B. So far, the problem is the same as ever: whatever one does, one will do what one does because of the way one's CPM is, and since one neither is nor can be ultimately responsible for the way one's CPM is, one can't be ultimately responsible for what one does.

Enter one's self, S. S is imagined to be in some way independent of one's CPM. S (i.e. one) considers the deliverances of one's CPM and decides in the light of them, but it—S—incorporates a power of decision that is independent of one's CPM 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 isn't ultimately responsible for one's CPM. The idea is that step (2) of the Basic Argument is false because of the existence of S (for a development of this view see, for example, Campbell 1957).

The trouble with the picture is obvious. S (i.e., one) decides on the basis of the deliverances of one's CPM. But whatever S decides, it decides as it does because of the way it is (or because of the occurrence in the decision process of indeterministic factors for which it—i.e. one—can't be responsible, and which can't plausibly be thought to contribute to its ultimate moral responsibility). And this brings us back to where we started. To be a source of ultimate responsibility, S must be responsible for being the way it is. But this is impossible, for the reasons given in the Basic Argument. So while the story of S and CPM adds another layer to the description of the human decision process, it can't change the fact that human beings cannot be ultimately self-determining in such a way as to be ultimately morally responsible for how they are, and thus for how they decide and act.

In spite of all these difficulties, many of us (nearly all of us) continue to believe that we are truly morally responsible agents in the strongest possible sense. Many of us, for example, feel that our capacity for fully explicit self-conscious deliberation in a situation of choice suffices—all by itself—to constitute us as such. All that is needed for true or ultimate responsibility, on this view, is that one is in the moment of action *fully self-consciously aware of oneself as an agent facing choices*. 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: 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, on this view, simply annihilated by one's self-conscious awareness of one's situation.

This is an extremely natural intuition; but the Basic Argument appears to show that it is a mistake. For however self-consciously aware we are as we deliberate and reason, every act and operation of our mind happens as it does as a result of features for which we are ultimately in no way responsible. And yet the conviction that self-conscious awareness of one's situation can be a sufficient foundation of strong free will is extremely powerful. It runs deeper than rational argument, and it survives untouched, in the everyday conduct of life, even after the validity of the Basic Argument has been admitted. Nor, probably, should we wish it otherwise.⁷

NOTES

1. Wouldn't it be enough if one simply endorsed the way one found oneself to be, mentally, in the relevant respects, without actually changing anything? Yes, if one were ultimately responsible for having the principles in the light of which one endorsed the way one found oneself to be. But how could this be?

2. See, e.g., MacKay (1960); Strawson (2010, 246–50; 1986, 281–86). When I cite a work I give the date of first publication, or occasionally the date of composition, while the page reference is to the edition listed in the bibliography.

3. Nussbaum has something much less dramatic in mind, I think, when she writes that “one’s own character is one’s own responsibility and not that of others” (Nussbaum 2004).

4. For some discussion of the deep ways in which we’re naturally compatibilist in our thinking about free will or moral responsibility, and don’t feel that it is threatened either by determinism or by our inability to be self-determining, see Strawson 1986, §6.4 (“Natural compatibilism”). Clarke and Fischer are prominent among those who misrepresent my position on free will to the extent that they focus only on the line of thought set out in the current paper. See, e.g., Clarke (2005), Fischer (2006b).

5. During the symposium on free will at the British Academy in July 2010, J. R. Lucas objected that this claim involved an equivocation. He suggested that it operated simultaneously as a conceptual claim and as a causal claim, in a way which vitiated it. I agree that it is both a conceptual claim and a causal claim, but not that this vitiates it. The following is a *conceptual* truth about the *causation* of intentional action: that with regard to the respect in which it is true to say that the action is intentional, it must be true to say that the agent does what he does—given, as always, the situation in which he finds himself or takes himself to be—because (this is a causal “because”) of the way he is; and indeed because of the way he is in some mental respect; whatever else is true, and whatever else may be going on. The truth of this claim is wholly compatible with the fact that the way you are when you act is a function of many things, including of course your experience of your situation—which is part of the way you are mentally speaking. Certainly the way you are mentally speaking isn’t just a matter of your overall character or personality, and the present argument has its full force even for those who question or reject the explanatory viability of the notion of character when it comes to the explanation of action (see, e.g., Harman (1999, 2000a); Doris (2002); see also Note 7 below).

6. Darwin (1838, 608). For “wicked” in the first line Darwin has “wrecked” (a characteristic slip).

7. On this last point, see, e.g., P. F. Strawson (1962); for a doubt, see Smilansky (1994). It will be interesting to see how the conviction of free will stands up to increasing public awareness of results in experimental and social psychology, which show that our actions are often strongly influenced by factors, situational or otherwise, of which we are completely unaware (see, e.g., Doris (2002), Wilson (2002), Nahmias (2007), Knobe and Nichols (2008)). The general effect of this “situationist” line of enquiry is to cast increasing doubt on our everyday picture of ordinary adult human agents as consciously aware of, and in control of, themselves and their motivations and subsequent actions in such a way that they are, generally speaking, fully morally responsible for what they do. Situationism finds a natural ally in Freudian theory, while considerably extending the range of factors that threaten to undermine our everyday picture of responsibility. It tells us that we are far more “puppets of circumstances” than we realize; it questions our conception of ordinary human beings as genuinely free agents in a way that is independent of any considerations about determinism or the impossibility of self-origination. At the same time (again in line with Freudian theory) it grounds a sense in which greater self-knowledge, a better understanding of what motivates one, can increase one’s control of and responsibility for one’s actions.