

The Impossibility of Moral Responsibility

Author(s): Galen Strawson

Reviewed work(s):

Source: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 75, No. 1/2, Free Will, Determinism, and Moral Responsibility (Aug., 1994), pp. 5-24

Published by: [Springer](#)

Stable URL: <http://www.jstor.org/stable/4320507>

Accessed: 20/08/2012 03:46

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Springer is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*.

THE IMPOSSIBILITY OF MORAL RESPONSIBILITY

(Received 15 September 1993)

I

There is an argument, which I will call the Basic Argument, which appears to prove that we cannot be truly or ultimately morally responsible for our actions. According to the Basic Argument, it makes no difference whether determinism is true or false. We cannot be truly or ultimately morally responsible for our actions in either case.

The Basic Argument has various expressions in the literature of free will, and its central idea can be quickly conveyed. (1) Nothing can be *causa sui* – nothing can be the cause of itself. (2) In order to be truly morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects. (3) Therefore nothing can be truly morally responsible.

In this paper I want to reconsider the Basic Argument, in the hope that anyone who thinks that we can be truly or ultimately morally responsible for our actions will be prepared to say exactly what is wrong with it. I think that the point that it has to make is obvious, and that it has been underrated in recent discussion of free will – perhaps because it admits of no answer. I suspect that it is obvious in such a way that insisting on it too much is likely to make it seem less obvious than it is, given the innate contrasuggestibility of human beings in general and philosophers in particular. But I am not worried about making it seem less obvious than it is so long as it gets adequate attention. As far as its validity is concerned, it can look after itself.

A more cumbersome statement of the Basic Argument goes as follows.¹

- (1) Interested in free action, we are particularly interested in actions that are 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 is also a function of one's height, one's strength, one's place and time, and so on. But the mental factors 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, mentally speaking, in certain respects, one must have brought it about that one is the way one is, mentally speaking, in certain respects. And it is not merely that one must have caused oneself to be the way one is, mentally speaking. One must have consciously and explicitly chosen to be the way one is, mentally speaking, in certain respects, and one must have succeeded in bringing it about that one is that way.
- (5) But one cannot really be said to choose, in a conscious, reasoned, fashion, to be the way one is mentally speaking, in any respect at all, unless one already exists, mentally speaking, already equipped with some principles of choice, 'P1' – preferences, values, pro-attitudes, 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, mentally speaking, in certain 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, i.e. (7), 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.²
- (10) So true moral responsibility is impossible, because it requires true self-determination, as noted in (3).

This may seem contrived, but essentially the same argument can be given in a more natural form. (1) It is undeniable that one is the way one is, initially, as a result of heredity and early experience, and it is undeniable that these are things for which one cannot be held to be in any responsible (morally or otherwise). (2) One cannot at any later stage of life hope to accede to true moral responsibility for the way one is by trying to change the way one already is as a result of heredity 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 at change, will be determined by how one already is as a result of heredity 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 heredity and previous experience. (5) This may not be the whole story, for it may be that some changes in the way one is are traceable not to heredity and experience but to the influence of indeterministic or random factors. But it is absurd to suppose that indeterministic or random factors, for which one is *ex hypothesi* in no way responsible, can in themselves contribute in any way to one's being truly morally 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 Europeans). The claim is only that people cannot be supposed to change themselves in such a way as to be or become truly or ultimately morally responsible for the way they are, and hence for their actions.

II

I have encountered two main reactions to the Basic Argument. On the one hand it convinces almost all the students with whom I have discussed the topic of free will and moral responsibility.³ On the other hand it often tends to be dismissed, in contemporary discussion of free will and moral responsibility, as wrong, or irrelevant, or fatuous, or too rapid, or an expression of metaphysical megalomania.

I think that the Basic Argument is certainly valid in showing that we cannot be morally responsible in the way that many suppose. And I think that it is the natural light, not fear, that has convinced the students I have taught that this is so. That is why it seems worthwhile to restate the argument in a slightly different – simpler and looser – version, and to ask again what is wrong with it.

Some may say that there is nothing wrong with it, but that it is not very interesting, and not very central to the free will debate. I doubt whether any non-philosopher or beginner in philosophy would agree with this view. If one wants to think about free will and moral responsibility, consideration of some version of the Basic Argument is an overwhelmingly natural place to start. It certainly has to be considered at some point in a full discussion of free will and moral responsibility, even if the point it has to make is obvious. Belief in the kind of absolute moral responsibility that it shows to be impossible has for a long time been central to the Western religious, moral, and cultural tradition, even if it is now slightly on the wane (a disputable view). It is a matter of historical fact that concern about moral responsibility has been the main motor – indeed the *ratio essendi* – of discussion of the issue of free will. The only way in which one might hope to show (1) that the Basic Argument was not central to the free will debate would be to show (2) that the issue of moral responsibility was not central to the free will debate. There are, obviously, ways of taking the word ‘free’ in which (2) can be maintained. But (2) is clearly false none the less.⁴

In saying that the notion of moral responsibility criticized by the Basic Argument is central to the Western tradition, I am not suggesting that it is some artificial and local Judaeo-Christian-Kantian construct that is found nowhere else in the history of the peoples of the world,

although even if it were that would hardly diminish its interest and importance for us. It is natural to suppose that Aristotle also subscribed to it,⁵ and it is significant that anthropologists have suggested that most human societies can be classified either as 'guilt cultures' or as 'shame cultures'. It is true that neither of these two fundamental moral emotions necessarily presupposes a conception of oneself as truly morally responsible for what one has done. But the fact that both are widespread does at least suggest that a conception of moral responsibility similar to our own is a natural part of the human moral-conceptual repertoire.

In fact the notion of moral responsibility connects more tightly with the notion of guilt than with the notion of shame. In many cultures shame can attach to one because of what some member of one's family – or government – has done, and not because of anything one has done oneself; and in such cases the feeling of shame need not (although it may) involve some obscure, irrational feeling that one is somehow responsible for the behaviour of one's family or government. The case of guilt is less clear. There is no doubt that people can feel guilty (or can believe that they feel guilty) about things for which they are not responsible, let alone morally responsible. But it is much less obvious that they can do this without any sense or belief that they are in fact responsible.

III

Such complications are typical of moral psychology, and they show that it is important to try to be precise about what sort of responsibility is under discussion. What sort of 'true' moral responsibility is being said to be both impossible and widely believed in?

An old story is very helpful in clarifying this question. This is the story of heaven and hell. As I understand it, true moral responsibility is responsibility of such a kind that, if we have it, then it *makes sense*, at least, 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, for one certainly does not have to believe in any version of the story of heaven and hell in

order to understand the notion of true moral responsibility that it is being used to illustrate. Nor does one have to believe in any version of the story of heaven and hell in order to believe in the existence of true moral responsibility. On the contrary: many atheists have believed in the existence of true moral responsibility. The story of heaven and hell is useful simply because it illustrates, in a peculiarly vivid way, the *kind* of absolute or ultimate accountability or responsibility that many have supposed themselves to have, and that many do still suppose themselves to have. It very clearly expresses its scope and force.

But one does not have to refer to religious faith in order to describe the sorts of everyday situation that are perhaps primarily influential in giving rise to our belief in true 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. 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 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. Even if you believe that determinism is true, and that you will in five minutes time be able to look back and say that what you did was determined, this does not seem to undermine your sense of the absoluteness and inescapability of your freedom, and of your moral responsibility for your choice. The same seems to be true even if you accept the validity of the Basic Argument stated in section I, which concludes that one cannot be in any way ultimately responsible for the way one is and decides. In both cases, it remains true that as one stands there, one's freedom and true moral responsibility seem obvious and absolute to one.

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 are the fundamental source of our inability to give up belief in true or ultimate moral responsibility. There are further questions to be asked about why human beings experience these situations of choice as they do. It is an interesting question whether any cognitively sophisticated, rational, self-conscious agent must experience situations of choice in this way.⁶

But they are the experiential rock on which the belief in true moral responsibility is founded.

IV

I will restate the Basic Argument. First, though, I will give some examples of people who have accepted that some sort of true or ultimate responsibility for the way one is is a necessary condition of true or ultimate moral responsibility for the way one acts, and who, certain that they are truly morally responsible for the way they act, have believed the condition to be fulfilled.⁷

E.H. Carr held that “normal adult human beings are morally responsible for their own personality”. Jean-Paul Sartre talked of “the choice that each man makes of his personality”, and held that “man is responsible for what he is”. In a later interview he judged that his earlier assertions about freedom were incautious; but he still held that “in the end one is always responsible for what is made of one” in some absolute sense. Kant described the position very clearly when he claimed 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 radical 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 of oneself as a person who . . . is his own originator”. John Patten, the current British Minister for Education, a Catholic apparently preoccupied by the idea of sin, has claimed that “it is . . . self-evident that as we grow up each individual chooses whether to be good or bad.” It seems clear enough that he sees such choice as sufficient to give us true moral responsibility of the heaven-and-hell variety.⁸

The rest of us are not usually so reflective, but it seems 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 quite a

delicate one, for we do not ordinarily suppose that we have gone through some sort of active process of self-determination at some particular past time. Nevertheless it seems accurate to say 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.

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 are not experienced as foreign, but are rather 'identified' with (it is a necessary truth that it is 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, somehow in control of and answerable for how one is (even, perhaps, for aspects of one's character that one does not like). Here, then, I suggest that we find, semi-dormant in common thought, an implicit recognition of the idea that true moral responsibility for what one does somehow involves responsibility for how one is. Ordinary thought is ready to move this way under pressure.

There is, however, another powerful tendency in ordinary thought to think that one can be truly morally responsible even if one's character is ultimately wholly non-self-determined – simply because one is fully self-consciously aware of oneself as an agent facing choices. I will return to this point later on.

v

Let me now 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 none the less.

- (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) What you intentionally do, given the circumstances in which you (believe you) find yourself, flows necessarily from how you are.

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, given the circumstances in which you (believe you) find yourself.

Comment. Once again the qualification about ‘certain mental respects’ is one I will take for granted. Obviously one is not responsible for one’s sex, one’s basic body pattern, one’s height, and so on. But if one were not responsible for anything about oneself, how one could 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.

So

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

But

- (3) You cannot be truly responsible for the way you are, so you cannot 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 is not. Suppose that

- (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 you and you alone 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. Nothing can be *causa sui* in the required way. Even if such causal ‘aseity’ is allowed to belong unintelligibly to God, it cannot be 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).

The rephrased argument is essentially exactly the same as before, although the first two steps are now more simply stated. It may seem pointless to repeat it, but the questions remain. Can the Basic Argument simply be dismissed? It is really of no importance in the discussion of free will and moral responsibility? (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 cannot 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 in spite of this argument, and 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 evident consequence of the Basic

Argument is that there is a fundamental sense in which no punishment or reward is ever ultimately 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) colour of their hair or the (natural) shape of their faces. The point seems obvious, and yet it contradicts a fundamental part of our natural self-conception, and there are elements in human thought that move very deeply against it. When it comes to questions of responsibility, we tend to feel that we are somehow responsible for the way we are. Even more importantly, perhaps, we tend to feel that our explicit self-conscious awareness of ourselves as agents who are able to deliberate about what to do, in situations of choice, suffices to constitute us as morally responsible free agents in the strongest sense, whatever the conclusion of the Basic Argument.

VI

I have 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 will briefly consider three sorts of response to the Basic Argument, and I will concentrate on their more simple expressions, in the belief that truth in philosophy, especially in areas of philosophy like the present one, is almost never very complicated.

(I) The first is *compatibilist*. Compatibilists believe that one can be a free and morally responsible agent even if determinism is true. Roughly, they claim, with many variations of detail, that one may correctly be said to be truly responsible for what one does, when one acts, just so long as one is not caused to act 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). Clearly, this sort of compatibilist responsibility does not require that one should be truly responsible for how one is in any way at all, and so step (2) of the Basic Argument comes out as false. One can have compatibilist responsibility even if the way one is is totally determined by factors entirely outside one's control.

It is for this reason, however, that compatibilist responsibility famously fails to amount to any sort of true *moral* responsibility, given the natural, strong understanding of the notion of true moral responsibility (characterized above by reference to the story of heaven and hell). One does what one does entirely because of the way one is, and one is in no way ultimately responsible for the way one is. So how can one be justly punished for anything one does? Compatibilists have given increasingly refined accounts of the circumstances in which punishment may be said to be appropriate or intrinsically fitting. But they can do nothing against this basic objection.

Many compatibilists have never supposed otherwise. They are happy to admit the point. They observe that the notions of true moral responsibility and justice that are employed in the objection cannot possibly have application to anything real, and suggest that the objection is therefore not worth considering. In response, proponents of the Basic Argument agree that the notions of true moral responsibility and justice in question cannot have application to anything real; but they make no apologies for considering them. They consider them because they are central to ordinary thought about moral responsibility and justice. So far as most people are concerned, they are the subject, if the subject is moral responsibility and justice.

(II) The second response is *libertarian*. Incompatibilists believe that freedom and moral responsibility are incompatible with determinism, and some of them are libertarians, who believe that that we are free and morally responsible agents, and that determinism is therefore false. In an ingenious statement of the incompatibilist-libertarian case, Robert Kane argues that agents in an undetermined world can have free will, for they can “have the power to make choices for which they have ultimate responsibility”. That is, they can “have the power to make choices which can only and finally be explained in terms of their own wills (i.e. character, motives, and efforts of will)”.⁹ Roughly, Kane sees this power as grounded in the possible occurrence, in agents, of efforts of will that have two main features: first, they are partly indeterministic in their nature, and hence indeterminate in their outcome; second, they occur in cases in which agents are trying to make a difficult choice between the options that their characters dispose them to consider. (The paradigm

cases will be cases in which they face a conflict between moral duty and non-moral desire.)

But the old objection to libertarianism recurs. How can this indeterminism help with *moral* responsibility? Granted that the truth of determinism rules out true moral responsibility, how can the falsity of determinism help? How can the occurrence of partly random or indeterministic events contribute in any way to one's being truly morally responsible either for one's actions or for one's character? If my efforts of will shape my character in an admirable way, and in so doing are partly indeterministic in nature, while also being shaped (as Kane grants) by my already existing character, why am I not merely lucky?

The general objection applies equally whether determinism is true or false, and can be restated as follows. We are born with a great many genetically determined predispositions for which we are not responsible. We are subject to many early influences for which we are not responsible. These decisively shape our characters, our motives, the general bent and strength of our capacity to make efforts of will. We may later engage in conscious and intentional shaping procedures – call them S-procedures – designed to affect and change our characters, motivational structure, and wills. Suppose we do. The question is then why we engage in the particular S-procedures that we do engage in, and why we engage in them in the particular way that we do. The general answer is that we engage in the particular S-procedures that we do engage in, given the circumstances in which we find ourselves, because of certain features of the way we already are. (Indeterministic factors may also play a part in what happens, but these will not help to make us responsible for what we do.) And these features of the way we already are – call them character features, or C-features – are either wholly the products of genetic or environmental influences, deterministic or random, for which we are not responsible, or are at least partly the result of earlier S-procedures, which are in turn either wholly the product of C-features for which we are not responsible, or are at least partly the product of still earlier S-procedures, which are in turn either the products of C-features for which we are not responsible, or the product of such C-features together with still earlier S-procedures – and so on. In the end, we reach the first S-procedure, and this will have been engaged in, and engaged in

the particular way in which it was engaged in, as a result of genetic or environmental factors, deterministic or random, for which we were not responsible.

Moving away from the possible role of indeterministic factors in character or personality formation, we can consider their possible role in particular instances of deliberation and decision. Here too it seems clear that indeterministic factors cannot, in influencing what happens, contribute to true moral responsibility in any way. In the end, whatever we do, we do it either as a result of random influences for which we are not responsible, or as a result of non-random influences for which we are not responsible, or as a result of influences for which we are proximally responsible but not ultimately responsible. The point seems obvious. Nothing can be ultimately *causa sui* in any respect at all. Even if God can be, we can't be.

Kane says little about moral responsibility in his paper, but his position seems to be that true moral responsibility is possible if indeterminism is true. It is possible because in cases of "moral, prudential and practical struggle we . . . are truly 'making ourselves' in such a way that we are ultimately responsible for the outcome". 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" (op. cit., p. 252). It is for this reason that we can be ultimately responsible and morally responsible not only in cases of struggle in which we are 'making ourselves', but also for choices and actions which do not involve struggle, flowing unopposed from our character and motives.

In claiming that we can be ultimately responsible for our present motives and character, Kane appears to *accept* step (2) of the Basic Argument. He appears to accept that we have to 'make ourselves', and so be ultimately responsible for ourselves, in order to be morally responsible for what we do.¹⁰ The problem with this suggestion is the old one. In Kane's view, a person's 'ultimate responsibility' for the outcome of an effort of will depends essentially on the partly indeterministic nature of the outcome. This is because it is only the element of indeterminism that prevents prior character and motives from fully explaining the outcome of the effort of will (op. cit, p. 236). But how

can this indeterminism help with moral responsibility? How can the fact that my effort of will is indeterministic in such a way that its outcome is indeterminate make me truly responsible for it, or even help to make me truly responsible for it? How can it help in any way at all with moral responsibility? How can it make punishment – or reward – ultimately just?

There is a further, familiar problem with the view that moral responsibility depends on indeterminism. If one accepts the view, one will have to grant that it is impossible to know whether any human being is ever morally responsible. For moral responsibility now depends on the falsity of determinism, and determinism is unfalsifiable. There is no more reason to think that determinism is false than that it is true, in spite of the impression sometimes given by scientists and popularizers of science.

(III) The third option begins by accepting that one cannot 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 cannot be held to be ultimately responsible for one's character or personality or motivational structure. This picture has some support in the 'phenomenology' of human choice – we sometimes experience our choices and decisions as if the picture were an accurate one. But it is easy to show that it cannot be accurate in such a way that we can be said to be truly or ultimately morally responsible for our choices or actions.

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 for both A and B. One's CPM makes one tend towards 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 cannot 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 is not ultimately responsible for one's CPM. Step (2) of the Basic Argument is false because of the existence of S.¹¹

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 else because partly or wholly because of the occurrence in the decision process of indeterministic factors for which it – i.e. one – cannot be responsible, and which cannot plausibly be thought to contribute to one's true moral responsibility). And this returns us to where we started. To be a source of true or ultimate responsibility, S must be responsible for being the way it is. But this is impossible, for the reasons given in the Basic Argument.

The story of S and CPM adds another layer to the description of the human decision process, but it cannot 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. The story is crudely presented, but it should suffice to make clear that no move of this sort can solve the problem.

'Character is destiny', as Novalis is often reported as saying.¹² The remark is inaccurate, because external circumstances are part of destiny, but the point is well taken when it comes to the question of moral responsibility. Nothing can be *causa sui*, and in order to be truly morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects. One cannot institute oneself in such a way that one can take over true or assume moral responsibility for how one is in such a way that one can indeed be truly morally responsible for what one does. This fact is not changed by the fact that we may be unable not to think of ourselves as truly morally responsible in ordinary circumstances. Nor is it changed by the fact that it may be a very good thing that we have this inability – so that we might

wish to take steps to preserve it, if it looked to be in danger of fading. As already remarked, many human beings are unable to resist the idea that it is their capacity for fully explicit self-conscious deliberation, in a situation of choice, that suffices to constitute them as truly morally responsible agents in the strongest possible sense. The Basic Argument shows that this is a mistake. 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. But the conviction that self-conscious awareness of one's situation can be a sufficient foundation of strong free will is very 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.

VII

There is nothing new in the somewhat incantatory argument of this paper. It restates certain points that may be in need of restatement. "Everything has been said before", said André Gide, echoing La Bruyère, "but since nobody listens we have to keep going back and beginning all over again." This is an exaggeration, but it may not be a gross exaggeration, so far as general observations about the human condition are concerned.

The present claim, in any case, is simply this: time would be saved, and a great deal of readily available clarity would be introduced into the discussion of the nature of moral responsibility, if the simple point that is established by the Basic Argument were more generally acknowledged and clearly stated. Nietzsche thought that thoroughgoing acknowledgement of the point was long overdue, and his belief that there might be moral advantages in such an acknowledgement may deserve further consideration.¹³

NOTES

¹ Adapted from G. Strawson, 1986, pp. 28–30.

² That is, the infinite series must have a beginning and an end, which is impossible.

³ Two have rejected it in fifteen years. Both had religious commitments, and argued, on general and radical sceptical grounds, that we can know almost nothing, and cannot therefore know that true moral responsibility is not possible in some way that we do not understand.

⁴ It is notable that both Robert Kane (1989) and Alfred Mele (forthcoming), in two of the best recent incompatibilist discussions of free will and autonomy, have relatively little to say about moral responsibility.

⁵ Cf. *Nichomachean Ethics* III. 5.

⁶ Cf. MacKay (1960), and the discussion of the ‘Genuine Incompatibilist Determinist’ in G. Strawson (1986, pp. 281–6).

⁷ I suspect that they have started out from their subjective certainty that they have true moral responsibility. They have then been led by reflection to the realization that they cannot really have such moral responsibility if they are not in some crucial way responsible for being the way they are. They have accordingly concluded that they are indeed responsible for being the way they are.

⁸ Carr in *What Is History?*, p. 89; Sartre in *Being and Nothingness, Existentialism and Humanism*, p. 29, and in the *New Left Review* 1969 (quoted in Wiggins, 1975); Kant in *Religion within the Limits of Reason Alone*, p. 40, *The Critique of Practical Reason*, p. 101 (Ak. V. 98), and in *Opus Postumum*, p. 213; Patten in *The Spectator*, January 1992.

These quotations raise many questions which I will not consider. It is often hard, for example, to be sure what Sartre is saying. But the occurrence of the quoted phrases is significant on any plausible interpretation of his views. As for Kant, it may be thought to be odd that he says what he does, in so far as he grounds the possibility of our freedom in our possession of an unknowable, non-temporal noumenal nature. It is, however, plausible to suppose that he thinks that radical or ultimate self-determination must take place even in the noumenal realm, in some unintelligibly non-temporal manner, if there is to be true moral responsibility.

⁹ Kane (1989) p. 254. I have omitted some italics.

¹⁰ He cites Van Inwagen (1989) in support of this view.

¹¹ Cf. C.A. Campbell (1957).

¹² e.g. by George Eliot in *The Mill on the Floss*, book 6, chapter 6. Novalis wrote “Oft fühl ich jetzt . . . [und] je tiefer einsehe, dass Schicksal und Gemüt Namen eines Begriffes sind” – “I often feel, and ever more deeply realize, that fate and character are the same concept”. He was echoing Heraclitus, Fragment 119 DK.

¹³ Cf. R. Schacht (1983) pp. 304–9. The idea that there might be moral advantages in the clear headed admission that true or ultimate moral responsibility is impossible has recently been developed in another way by Saul Smilansky (1994).

REFERENCES

- Aristotle, 1953. *Nicomachean Ethics*, trans. J. A. K. Thomson, Allen and Unwin, London.
- Campbell, C.A., 1957. 'Has the Self "Free Will"?', in C.A. Campbell, *On Selfhood and Godhood*, Allen and Unwin, London.
- Carr, E.H., 1961. *What Is History?*, Macmillan, London.
- Kane, R., 1989. 'Two Kinds of Incompatibilism', *Philosophy and Phenomenological Research* 50, pp. 219–254.
- Kant, I., 1956. *Critique of Practical Reason*, trans. L. W. Beck, Bobbs-Merrill, Indianapolis.
- Kant, I., 1960. *Religion within the Limits of Reason Alone*, trans. T. M. Greene and H. H. Hudson, Harper and Row, New York.
- Kant, I., 1993. *Opus postumum*, trans. E. Förster and M. Rosen, Cambridge University Press, Cambridge.
- MacKay, D.M., 1960. 'On the Logical Indeterminacy of Free Choice', *Mind* 69, pp. 31–40.
- Mele, A., 1995. *Autonomous Agents: From Self-Control to Autonomy*, Oxford University Press, New York.
- Nietzsche, F., 1966. *Beyond Good and Evil*, trans. Walter Kaufmann, Random House, New York.
- Novalis, 1802. *Heinrich von Ofterdingen*.
- Sartre, J.-P., 1969. *Being and Nothingness*, trans. Hazel E. Barnes, Methuen, London.
- Sartre, J.-P., 1989. *Existentialism and Humanism*, trans. Philip Mairet, Methuen, London.
- Schacht, R., 1983. *Nietzsche*, Routledge and Kegan Paul, London.
- Smilansky, S., 1994. 'The Ethical Advantages of Hard Determinism', *Philosophy and Phenomenological Research*.
- Strawson, G., 1986. *Freedom and Belief*, Clarendon Press, Oxford.
- Van Inwagen, P., 1989. 'When Is the Will Free?', *Philosophical Perspectives* 3, pp. 399–422.
- Wiggins, D., 1975. 'Towards a Reasonable Libertarianism', in T. Honderich, ed., *Essays on Freedom of Action*, Routledge, London.

Jesus College, Oxford
Oxford OX1 3DW
England