

Cambridge University Press

0521472792 - Making Sense of Humanity: And Other Philosophical Papers 1982-1993

Bernard Williams

Excerpt

[More information](#)

---

I

Action, freedom, responsibility

---

Cambridge University Press

0521472792 - Making Sense of Humanity: And Other Philosophical Papers 1982-1993

Bernard Williams

Excerpt

[More information](#)

---

1

---

## How free does the will need to be?

---

### 1 The irrelevance of constraint

Locke said<sup>1</sup> that the question was not whether the will be free, but whether we have a will. *A fortiori*, it cannot be a question of how free the will may be. Locke's remark reminds us that the freedom of the will that has been the subject of the classical problem, if it comes at all, does not come in degrees. People's freedom, however, in more everyday senses, obviously enough does come in degrees.

This point raises a question not only about the classical problem of freewill, but also about the classical compatibilist answer to it, based on the idea that freedom is opposed to constraint and not to necessity. That position certainly deploys an idea of freedom, but of a freedom that may be more or less extensive, and that fact in itself should make us ask whether the position does not miss the point of the problem it is supposed to resolve. It is far from clear what exactly constraint is,<sup>2</sup> but in the kinds of cases usually invoked, somebody brings it about, by intentional application of threats or force, that an agent either cannot attain his original goal at all, or can attain it only at an increased cost. The agent may be confronted by a highwayman who (whatever the standard formula says) does not, in fact, offer him a choice between his money and his life, but rather a choice between losing merely his money, and losing his money together with his life; in that case, he cannot attain the goal of keeping his money at all. On the other hand, a man who possesses a valuable secret may be able to attain at least part of his objectives, and frustrate his captors, if their threat to kill him leaves him with the option of dying with his secret.

But what significance is there in the fact that the cases standardly invoked are cases of *constraint* – that is to say, cases in which the

Cambridge University Press

0521472792 - Making Sense of Humanity: And Other Philosophical Papers 1982-1993

Bernard Williams

Excerpt

[More information](#)*Action, freedom, responsibility*

limitation of effective choice is deliberately imposed, for their own ends, by other agents? These are merely one sort of what Aristotle rightly and relevantly identified as 'actions done through fear of greater evils', such as that of the sailors who throw the cargo overboard to save themselves and the ship.<sup>3</sup> In the cases that do not involve other people's hostile intentions, the agent's original objective may equally be made more costly, or it may become inaccessible, or his objectives may need to be modified in one way or another.

Now in all cases of things done through fear of greater evils, whether that fear is imposed by other agents or not, there is no loss of *freewill* in any sense that has to do with the agent's capacity to choose, or with his being held responsible. The agent is responsible for his action. He may not be to blame for the loss of the goods or whatever, in the sense that a course of action that would have been unreasonable or blameworthy in ordinary circumstances is reasonable and not blameworthy in the constrained circumstances. This is very obvious when the agent's original objective is accessible, though at greater cost, but similar points apply to the cases in which a course of action usually or previously thought available becomes unavailable. People are indeed seen as responsible for their actions in such circumstances, as when Aristotle's sailors are complimented for their prudence, or alternatively told off for panicking. In such circumstances you can of course be excused for not doing something that you would otherwise be blamed for not doing; more than that, you can be complimented for not trying to do it.

If we look at the larger class of things done through fear of greater evils, we are not going to learn much relevant to freewill or to ideas of compatibilism. It might be said that these are things that people do 'against their will'. But things done against one's will, in such circumstances, are not even (except in a very everyday sense) things that one does when not wanting to do them. They are things that one did not originally want to do, or which one would not want to do unless the circumstances were exceptionally disadvantageous, and their possibility does not shed any more light on the question of compatibilism than that which is shed by other actions that are performed under limiting circumstances.

If we are not to count as exercising freewill in cases of this kind, then we never exercise it, since all choices operate in a space of alternatives constrained by the contingent cost of various possibilities, and these exceptional cases are simply dramatic cases of that, where the space has been unexpectedly restricted. It makes no difference to this central aspect, so far as the agent's decisions and their status are concerned,

Cambridge University Press

0521472792 - Making Sense of Humanity: And Other Philosophical Papers 1982-1993

Bernard Williams

Excerpt

[More information](#)*How free does the will need to be?*

whether the space of possibilities has been altered by a human being with the intention of doing just that. But the cases in which that is so are the ones that count as cases of constraint. So constraint is a red herring so far as freewill is concerned.

There are of course some important differences between actions done under constraint, and other actions done from fear of greater evils. If the restrictions are humanly imposed, they are likely to elicit resentment as well as frustration. Moreover, constraint is peculiarly related to the deliberative conclusion that one *must* or *has to* do a certain thing (a kind of conclusion to which I shall return). Conclusions reached under constraint are not always of that form, but they often are. When such conclusions are reached in other situations, they characteristically express some project or objective with which the agent is deeply identified, for instance (though by no means exclusively) requirements of morality. What is peculiar about these conclusions, when they are reached under constraint, is that they witness to the agent's vital interests or deepest needs only negatively, as things to be protected: the actions required are the expression of someone else's intentions and can lie indefinitely far away from anything with which the agent is identified.

Such considerations can indeed help to explain why constraint – restrictions imposed by other people, rather than by nature – is perceived as specially opposed to *freedom*. It constitutes being in someone else's power, which is a paradigm of unfreedom. But the very fact that decisions taken under constraint are decisions – and that, further, they can take the form of practical necessity, a form that belongs to some of the most serious and responsible decisions we take – shows why constraint cannot provide the clue to understanding freewill.

Why does freewill, unlike freedom, not come in degrees? Presumably it is because its assertion consists only of an existence claim. How exactly that claim should be expressed is notoriously disputed, but it is something to the effect that agents sometimes act voluntarily, and that when they do so they have a real choice between more than one course of action; or more than one course is open to them; or it is up to them which of several actions they perform. I shall leave the claim in this indeterminate form, and give it a label that preserves its indeterminacy, *the Plurality Principle*. Why the Principle takes the form of an existence claim will become clearer later. The present point is that it merely requires that there be, in the appropriate sense, alternatives for the agent, and that it is indifferent to their number, their cost, and so forth. That is why the freewill that it introduces is different from the freedom that comes in degrees and is opposed to constraint.

Cambridge University Press

0521472792 - Making Sense of Humanity: And Other Philosophical Papers 1982-1993

Bernard Williams

Excerpt

[More information](#)*Action, freedom, responsibility*

## 2 What the reconciling project must be

The old compatibilism made a lot out of the opposition between freedom and constraint. If the argument of section 1 is right, it was looking in the wrong direction. But that is not the only reason why we have to recast the question of compatibilism, or, as we may say, the reconciling project. As it is usually described, the reconciling project involves an important, structural, misconception. Its task is explained in terms of taking two recognizable items – determinism (or something like that), on the one hand, and on the other hand, something that is often called ‘moral responsibility’ – and trying to reconcile them with one another. But this account of the task underdescribes it, because there are not two but three items or sets of items to be accounted for. They are, first, determinism (or something like that); second, a class of psychological items such as choice, decision, or rational action; and, third, some ethical items such as blame or responsibility. Since there are three items or classes of items involved, there is more than one way of understanding what would be involved in reconciliation. It may be thought that what need to be reconciled are determinism and choice, where choice is understood as a psychological item, and that if this can be achieved, the ethical notions will be able to live with determinism. Alternatively, it may be thought that even if the psychological items can be reconciled with determinism, this may not be enough to save the ethical notions, which require something more; something that excludes determinism. If this further demand is put in terms of *choice*, it might be expressed by saying that responsibility and similar ethical notions require real choice, and real choice is not a purely psychological notion, but a metaphysical one. I shall come back to this idea.

As well as undercounting the items involved, reconcilers have tended to make the further mistake of thinking that we understand the ethical items at least as well as we understand the psychological items. A similar mistake is often made by their opponents as well. Of course, the reconcilers and their opponents do not necessarily think that they have the *same* determinate understanding of the ethical notions; notions such as, in particular, ‘moral responsibility’. Because of these differences, the opponents sometimes say that the reconcilers have failed to show that our actual ethical notions are compatible with determinism, but, rather, have changed the subject and brought in a reductive and inadequate version of those notions. Both parties, however, do tend to agree on two things: that we have a determinate understanding of the relevant ethical notions; and that what we have to worry about, if anything, is just the relation of those notions to determinism (or something like that).

The truth is that we have other reasons to worry about many of our

Cambridge University Press

0521472792 - Making Sense of Humanity: And Other Philosophical Papers 1982-1993

Bernard Williams

Excerpt

[More information](#)*How free does the will need to be?*

moral notions, and if we have come to have difficulty in understanding ideas such as 'moral responsibility', this is not simply because of our suppositions, hopes and fears about naturalistic explanations of action. It is to some extent because of this, and inasmuch as compatibilism was, like its opponents, wedded to 'moral responsibility' as the ethical term that had to be reconciled to naturalistic explanation, it has failed. But, more significantly, we have reasons *anyway* for being doubtful about 'moral responsibility'.

I said that the first item to be considered in relation to the reconciling project was 'determinism' (or something like that). The important feature of this item is that it should be some general doctrine about the world's workings, which is thought, unless the reconciling project is successful, to conflict with the Plurality Principle. It is this item, in fact, that determines the shape that the Plurality Principle takes. The feared effect of the doctrine in question is to reduce in every case the actions open to the agent in the relevant sense to one, and it is this that the Principle denies.

It is much less important that the doctrine in question should be universal determinism. There may have been a time when belief in a universal determinism looked like the best reason there was for expecting strong naturalistic explanations of psychological states and happenings, but, if that was once the case, it is no longer so. It now looks a great deal more plausible and intelligible that there should be such explanations than that the universe should be a deterministic system, and it is the possibility of those explanations that itself creates the problem. 'Strong naturalistic explanation' is an extremely vague phrase, and it may be said that a good deal more needs to be known about what it means, before we can know what the problem is supposed to be. It may be said, in particular, that only tight nomological explanations can generate the problem. I doubt that this is correct. As Daniel Dennett<sup>4</sup> has pointed out, the chance of being able to cash in the nomological claim at the only level to which it could apply – that of a repeatable microstate – is vanishingly small. But the mere failure to do so (because of randomness, for instance) would be uninteresting in itself, while strong psychophysical explanations that did not meet that standard could equally create unease for the Plurality Principle, if anything of that sort could.

This follows from the nature of the Plurality Principle itself, which is expressed in terms of 'courses of action'. How many courses of action are 'open to' the agent (or whatever the preferred formulation may be) must be considered in terms of what can be counted, in relation to a given agent on a given occasion, as relevantly different courses of action. There is a complex question of how we should understand this, but it seems to be a sufficient condition of an agent's enjoying on a given occasion the

Cambridge University Press

0521472792 - Making Sense of Humanity: And Other Philosophical Papers 1982-1993

Bernard Williams

Excerpt

[More information](#)*Action, freedom, responsibility*

freewill defined by the Principle that the courses of action *which present themselves to that agent as deliberative alternatives* are in the required sense open to him. (Some of the complexities lie in the fact that this is not also a necessary condition: there may be various courses of action relevantly open to him, but he may have overlooked or misidentified them in his deliberation. But these problems need not be pursued in the present argument.)

Suppose that the Plurality promised by the Principle, so understood, fails: only one of the courses of action that present themselves to the agent in his choice is actually open to him. It is compatible with this that various different movements might still be possible, any of which would constitute this course of action, and that no microdeterministic set-up determined which set of movements would come about. This would not restore the required Plurality. On the other hand, the reason why Plurality failed might be found in strong psychophysical explanations operating at a level, perhaps, at which actions were specified. If so, it will be the strong explanations that matter, rather than microdeterminism.

It certainly remains unclear what the strongest kind of psychophysical explanations might be like. But I do not think that so far as these questions are concerned, it matters a great deal what exactly they might be like. It must in any case be sensible to test the psychological and ethical notions against the strongest hypotheses we could possibly entertain about psychophysical explanation. Further, there is a substantive reason for this policy. So far as these issues are concerned, the answers to questions about psychophysical explanation will matter only if the outlook for the psychological items is sensitive to those answers. But one respect, as it seems to me, in which the reconciling project has been a success is that the outlook for the psychological items has been shown not to be sensitive to those answers. Work by Brian O'Shaughnessy<sup>5</sup> and others seems to me to have shown that those concepts can function compatibly with the strongest hypotheses about explanations. We have good reason to believe that, insofar as they are psychological notions, all the following are compatible with any conceivable possibility of naturalistic explanation: choice; reasoned choice, and decision; action; intentional action; reasoned intentional action; and what is entailed by that, trying.

### 3 Determinism, fatalism, and an incompatibilist argument

If this work has achieved the reconciling project so far as the psychological items are concerned, the remaining task should consist of bringing the ethical items into an intelligible relation to the psychological items.



Cambridge University Press

0521472792 - Making Sense of Humanity: And Other Philosophical Papers 1982-1993

Bernard Williams

Excerpt

[More information](#)*How free does the will need to be?*

However, opponents of the reconciling project – call them ‘incompatibilists’ – may say that to go about things in this way is to beg the question. They may say that it is a mistake to suppose that we can do with a purely ‘psychological’ conception of choice, or (perhaps) of intentional action, if that means a concept lacking the kinds of metaphysical implications that are relevant to this problem. They may say that the concepts we have do have such implications, and if a ‘psychological’ concept of choice or intentional action lacks them, then it is not an adequate concept. Alternatively, the incompatibilist may say that there is perhaps a purely psychological concept of choice under which choices or deliberated action can be reconciled with strong psychophysical explanation, but this concept is inadequate to our ethical purposes, which do require a concept with stronger metaphysical implications.

Sometimes it is hard to tell which of these views an incompatibilist is offering. This is particularly likely to be so when the concept under discussion is ‘free action’. This is the concept standardly associated with the Plurality Principle, but what work exactly is this concept expected to do? It is most often mentioned in connection with ethical demands, especially the demands of blame, and this might imply that ‘free action’ is above all an ethical concept. On this showing, if the Plurality Principle fails for metaphysical reasons, we may have to revise our ethical practices, but we could still be left with functioning psychological concepts of choice and intentional action. Some incompatibilists, however, seem to proceed on the assumption that, even apart from ethical considerations, ‘free action’ represents a basic category of action: much the same, perhaps, as ‘chosen action’ or ‘deliberated action’ or ‘intentional action’. It may even be claimed that it is coextensive with action itself, at least when that is understood in the strictest sense. On this account, the metaphysical implications of the Plurality Principle are such as to leave the whole conceptual structure of action in ruins if they turned out to be false.

This more radical incompatibilist line gives rise to a rather uneasy dialectical situation. Suppose that the incompatibilist argues that deliberated action (say) is incompatible with determinism (or something like that: for simplicity, I shall take the item to be determinism). It may seem that it is absolutely obvious that there is such a thing as deliberated action – because, for instance, we all go in for it. Then, if the incompatibilist argument is correct – more particularly, if it can be made obvious that it is correct – it will be obvious that determinism is false. But how could that be? You would suppose that if determinism were coherent at all, it would be a large task to establish its truth or falsehood. Faced with the obvious existence of deliberated action and the unobviousness of the truth or



*Action, freedom, responsibility*

falsehood of determinism, the most economical assumption is that there is something wrong with this style of incompatibilist argument: for instance, the incompatibilist, for ultimately ethical reasons, has interpreted the familiar concept of deliberated action in too ambitious a way. (It can be helpful to bear in mind the truism that if determinism is true at all, it is true *already*.) It is a product of this unstable situation that many incompatibilist arguments, including some of those most familiar from the tradition, prove (if anything) more than their authors can want. They turn out, for instance, not to distinguish satisfactorily between determinism and fatalism, even if the incompatibilist wants, reasonably, to be able to make this distinction.

This is illustrated by the version of a familiar incompatibilist argument that has been presented by David Wiggins.<sup>6</sup> Wiggins starts from the idea of what an agent can or cannot bring about at a given time, where this involves the idea of the steps that an agent may or may not be able to take at one time in order to bring about an outcome at a later time. It is uncontentionally illustrated by a situation in which at 9.55, being where I am, I may (still) be able to catch the train that leaves at 10.15, but at 10.03 I am no longer able to do so. In terms of this idea of a tensed capacity, Wiggins defines a notion of *historical inevitability at t*:

By 'it is historically inevitable at time *t'* that *p*' is intended something like this: whatever anybody does at *t'* or thereafter, it cannot make any difference to whether *p*...<sup>7</sup>

Let 'Nec<sub>*t'*</sub>' represent 'it is historically inevitable at *t'* that ...'. Then, under determinism, for each action, *R*, of an agent *A*, there will be some truth of the form

- (1) Nec<sub>*t'*</sub> (if *C* at *t* then *A* does *R* at *t'*),  
 which is a consequence of some law of nature, and expresses an inevitability because no-one can change the laws of nature. Moreover, for the appropriate set of conditions, *C*,
- (2) Nec<sub>*t'*</sub> (*C* at *t*),  
 simply because *C* obtained at *t*, and no-one can change the past. So if we take 'Nec<sub>*t'*</sub>' indeed to represent a modality, we can use 'the uncontroversial modal principle'
- (3) If Nec(*p*), then [if Nec(if *p* then *q*), then Nec(*q*)]  
 to reach the conclusion
- (4) Nec<sub>*t'*</sub> (*A* does *R* at *t'*),  
 and 'how', Wiggins asks, 'does this consist with'

Cambridge University Press

0521472792 - Making Sense of Humanity: And Other Philosophical Papers 1982-1993

Bernard Williams

Excerpt

[More information](#)*How free does the will need to be?*(5) A could at  $t'$  have done something other than R?<sup>8</sup>

How indeed, since by the definition of inevitability (4) entails

(6) Whatever A does at  $t'$ , A does R at  $t'$ 

and (6), on a natural reading, expresses a contradiction. This may just be a failing of the definition of 'historical inevitability'. But what interpretation of (4) will give the result that the argument needs? I cannot see any, constructed from these materials, that will not express a fatalistic conclusion.

Wiggins denies that his argument involves fatalism, but his ground for this is simply that it involves contingent premises, and is not 'a mere logical puzzle'.<sup>9</sup> But this is to reduce fatalism to the type that has been called 'logical fatalism'. 'Fatalistic' does not pick out a class of arguments, but a class of beliefs or attitudes: those that involve the idea that action, choice, and so forth have no effects; that everything will be the same whatever you do.<sup>10</sup> In many cases, such a belief is manifestly false, which is the main reason why incompatibilists want to keep determinism and fatalism apart: if determinism, by their argument, is as manifestly false as fatalism, then, as I just suggested, this casts doubt on their argument. With Wiggins's argument, however, the attempt to keep fatalism and determinism apart has failed.

This comes out if we reflect that, on the basis of the directions and the examples we have been given, we can read ' $\text{Nec}_t(p)$ ' as saying 'no thought or intention of any agent, occurring at  $t$  or later, has an effect on the obtaining of the state of affairs that makes  $p$  true'. In this sense of the operator, not only are (1) and (2) true, but so are

(1!)  $\text{Nec}_t$  (if C at  $t$  then A does R at  $t'$ )

and

(2!)  $\text{Nec}_t$  (C at  $t$ ),

so by the modal principle we can derive

(4!)  $\text{Nec}_t$  (A does R at  $t'$ ),where  $t$  can be an arbitrarily long time before  $t'$ .

(4!), together with the explanation of the operator, entails that no thought or intention of A in an arbitrarily long period before  $t'$  has an effect on what A does at  $t'$ , and this is a nakedly fatalistic conclusion.

Unless one wants to argue that determinism simply does entail an absurd fatalistic conclusion, it is reasonable to conclude that there is something wrong with this argument. It seems fairly clear what is wrong.