Alternate Possibilities and their Entertainment

S. ROUSH

Harry Frankfurt put us on notice that the principle of alternate possibilities—'a person is morally responsible for what he has done only if he could have done otherwise'—, a principle that was presumed without argument in much discussion of free will and determinism, is not true. He proposed to replace it with the claim that 'a person is not morally responsible for what he has done if he did it *only because* he could not have done otherwise,' a principle which 'does not appear to conflict with the view that moral responsibility is compatible with determinism.'[1] While Frankfurt's revised principle does have the merits he thinks it has with regard to representing our notion of responsibility more accurately, I will argue that it does not help us decide whether free will is compatible with determinism. We should be confident that free will and determinism are compatible only when we are sure we are not being evasive about the matter. We should be confident we are not being evasive when what produces the sense of conflict is aptly located. This is the question I hope to shed some light on here. In the process I will explain why the question about alternate possibilities and moral responsibility is relevant to Strawson's defence of compatibilism in his paper 'Freedom and Resentment', and why his defence is therefore insufficient. Finally, I will present an argument supporting Strawson's claim that we could not believe in a determinism that implied we should suspend all reactive attitudes toward all people, but conclude that if this argument is correct then it supports, if anything, *in*compatibilism.

Frankfurt's argument that the principle of alternate possibilities is incorrect appears sound, except precisely when the lesson of it is to be applied in the context of tensions between free will and determinism. The principle is incorrect because of the existence of a counter-example (or class of counter-examples) in whose vivid form a person called Black is poised to intervene if Jones is ever about to

[1] Harry Frankfurt, 'Alternate possibilities and moral responsibility', *The Importance of What We Care About* (Cambridge University Press, 1988), 1–10, on 1 and 10, italics mine. This essay was originally published in *Journal of Philosophy*, 1969, 829–39.

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

act otherwise than Black wishes.[2] Black is assumed capable through his planned interventions, of making Jones do exactly as he wishes if Jones does not decide to do it on his own.[3] In the case of interest, Jones never decides on his own to do otherwise than Black wishes him to, so Black never needs to intervene. Thus it is true both that Jones does everything he does of his own accord, and that he could not have done otherwise. Had he tried to do otherwise, Black wouldn't have let him. We would hold Jones responsible for his actions, since in fact he did them entirely of his own accord. Yet, since he could not have done otherwise, we have a case where we hold someone responsible for actions he could not have done otherwise, falsifying the principle of alternate possibilities.

The strength of this point is its weakness, however, when it comes to considering whether free will is compatible with determinism. While the counter-example shows that our notion of responsibility can be separated from that of the availability of alternate actions, and therefore makes a valuable conceptual point, it does not show that these things *are* separated in the kind of case at issue in worries about free will and determinism. In the counter-example, what insures that Jones cannot do otherwise than he does is a different thing from what makes him do things of his own accord. The first is Black, the second is Jones's 'free will' or some such. Moreover, it is only when the second is about not to do what the first desires that the first has any actual efficacy. Otherwise Black's role is only to insure the lack of alternate possibilities. This doesn't change if we substitute a computer for Black, or leave all other agents out of it by substituting natural forces which happen to work just as Black does. In any such case there are two entities, one which has reasons and can tell you why he did such and such, and the other on standby to act on this agent as if he were an object, in case he makes the 'wrong' decision. If Black, or the natural forces, were to act on Jones, it is unlikely that Jones the agent would feel he had done what he did for reasons of his own.

Therefore it appears that we have two entities, only one of which is actually efficacious, or even feels himself efficacious, at any given time, but one of which—Black, or his stand-in—governs the possible. Though there may be intermediate cases where it is true both that someone wanted to do something and that they were coerced or necessitated, and that both aspects had some actual efficacy in

[2] An example of the same general type has been proposed as a counter-example to the principles of alternate possibilities by Robert Nozick in lectures on free will given in 1966–67.

[3] This capability of Black's can be assumed to be present in whatever form the defender of the principle of alternate possibilities's understanding of 'can' requires.

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

bringing about the action, this mixed case will not do for falsifying the principle of alternate possibilities. This is because the extent to which the person was actually coerced—did it 'because of' the coercion—is the extent to which we will excuse her, but is also precisely the extent to which there were (actually) no alternate possibilities. To falsify the principle, it appears we need an agent fully responsible for what actually happens because bringing it about, and something else governing what is possible though only standing by 'just in case.'

The difficulty with this as a contribution to the question of whether there is a tension between free will and determinism is that this kind of situation is hardly a fit description of the case we are dealing with there. Whatever the thesis of determinism is taken to be it is surely not the idea that the laws of nature, or natural causes, are standing by 'just in case' we attempt to do otherwise than they dictate. And if we feel a problem about free will and determinism at all, it surely has something to do with the fact that the acts we perform, steering a car, say, are the same as physical events which would be determined, or caused or dictated, if determinism were true. We describe the acts of my steering the car differently than we describe my hands exerting certain forces on the steering wheel in certain directions, yet my acts would be nothing at all if they were not causally efficacious in the physical world in some way. In this sense they are to be identified with, if also described differently from, the physical events which are my steering of the car. There may be morally relevant acts which do not have causal counterparts or consequences in the physical world, yet there are enough which do that we can let the problem about free will and determinism be about those only. Morally relevant actions are very often also physical events. And it is because the very same action which I take myself to do of my own accord is also an event in a sequence of physical causes that the thesis of universal determinism gives us pause, when it does.

Since the counter-example scenario separates these two sides of Jones's actions, and makes only one side ever *actually* efficacious, the point made with it sidesteps what I think is the difficult type of case that creates the feeling there is a tension between free will and determinism. We can agree with Frankfurt, for example, that the fact that an irresistible threat is made does not entail that the person who received it was coerced to do what he did; the threat must also be what actually accounts for his doing the act. Yet this is quite far from the sort of case which gives rise to a feeling that free will and determinism are incompatible. The difficult type of case is hard to describe or imagine: it is as if *both* Jones were doing things of his

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

own accord *and* Black were actually performing his machinations to bring those same actions about. Both, that is, need to be efficacious, and sufficient to bring the action about, if we are to take compatibilism (and incompatibilism) seriously. If we don't know how to imagine, or don't know what to say about, this case then I think we have located the genuine problem about free will and determinism. There are more than one, and therefore too many, sufficient reasons for, or accounts of, those actions of Jones. We recognize a problem about free will and determinism if we have the intuition that one of them feels otiose when the other is taken seriously. Thus Frankfurt's separation of the two sources for Jones's actions, and his allowing only one ever actually to be efficacious at a given time, introduce conditions which do not allow the paradigm case of incompatibility to arise.[4]

Frankfurt admits that if one revised the principle of alternate possibilities only by requiring for excuse from responsibility that it be *because* of the lack of alternatives that the person did what he did—the lesson learned from the counter-example—, then the use to which the principle has been put to maintain that responsibility is incompatible with determinism would be unaffected. This is because 'if it was causally determined that a person perform a certain action, then it will be true that the person performed it because of those causal determinants.'[5] It is then by consulting our actual practice of excusing, as he views it, that he revises the principle further, to say that a person is not morally responsible for an action if she did it *only* because she could not have done otherwise. This, I take it, is meant to allow for compatibility between free will and determinism because we are now able to see it as possible that a person did what she did *because* she was causally determined to do so, but at the same time *wanted* to do what she did, and therefore because her action wasn't done *only* because she was causally determined she can be held morally responsible.

But why is this plausible? The only concrete example we have been given of a person we agreed to hold responsible despite his being unable to do otherwise was Jones, and there it was because in addition to the fact that Jones did what he did of his own accord, what made it the case that he could not do otherwise (Black or his stand-in) was not actually efficacious in those actions; it only *might*

[4] In another way Frankfurt implicitly admits the deeper problem as I've described it, since in discussing various cases of coercion he takes it that we must choose whether the person's action was done because of coercion or because of their choice and, apparently, that the extent to which one is efficacious is the extent to which the other isn't.

[5] Frankfurt, 9.

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

have been. If it is causal determinism that is making it the case that a person cannot do otherwise than she does, then we have quite a different case, as Frankfurt admits, for then the action is happening *because* of the causal determinism. Moreover, if the causal determinism makes it the case that she cannot do otherwise, then it is sufficient to bring her action about. We are meant to hold her responsible because in addition to all of this the action was something she 'really wanted to do.' Thus, because she also had reasons for doing what she did, or at least wanted to do it, it was not *only* because of causal determinism that she did it. But this is lame; what Frankfurt misses is the sense the example of causal determinism introduces that it might as well have been *only* because of causal determinism that she did what she did, since the causal determinism was both sufficient to bring her action about and is admitted to have actually brought it about. Her agency looks otiose, her feeling that she really wanted to do it a mere epiphenomenon.

Before explaining this further let me point out something which the Jones example and the example of someone causally determined—call her 'Jane'—have in common. This is that neither is a case which our concept of responsibility has grown up coping with. This should be obvious in the case of Jones, but is also true with Jane. We do not in the normal run of affairs judge for responsibility cases where we believe both that an action was actually brought about by forces sufficient to make a person do what she did *and* that the person's wanting to do it was sufficient to bring the action about and actually brought the action about. This casts doubt on Frankfurt's appeal to our actual practice of excusing in order to decide the business about Jane. He thinks that when a person offers as an excuse the fact that he could not have done otherwise, the reason we excuse him if we do is that we understand by this more than the sentence conveys. We understand two more things: that he did it *only* because he could not have done otherwise, and he did not perform the act because he wanted to do it. Let's suppose that this is true, as is plausible. The only reason it seems to me Frankfurt has given for holding Jane responsible is that she cannot similarly say it did not happen because she wanted to do what she did, and this is because she really wanted to do what she did. But to take this as deciding the issue of her responsibility or not is to take her case to be the same as the familiar one, which in an important respect it is not.

It is not the same because we do not in the normal case expect that an action was done both because of forces beyond the agent's control *and* because the agent wanted to do just that. The 'because' in each clause is important here. If one or the other of these factors

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

is both a sufficient reason for the act to have occurred and the action actually occurred for that reason, it crowds the other reason out and makes it seem otiose. This, I think, is in fact *why* what Frankfurt cites about our practices of excusing is the case. When we excuse someone for an action which he says happened *because* he could not have done otherwise, we understand more by this statement because we assume that a sufficient reason which actually causes the action allows for no other efficacious factors. We conclude that he means he did it only because he could not have done otherwise, and not because that's what he wanted to do, at least largely because of this crowding out phenomenon. A situation where an action is done both *because* of forces beyond a person's control and *because* she wanted to do it, where both factors are sufficient to bring the action about, is simply *not normal*.

We can ask, about the normal case, whether we excuse someone for an action that happened because the person could not have done otherwise because of the presence of an efficacious sufficient reason which was not his own will or because of the absence of his own will's efficacy. Frankfurt has assumed the latter, and this is why Jane's admission that she wanted to do what she did is supposed to make her responsible. I have claimed that in the normal case the former more or less automatically excludes the efficacy of the person's will, and that we excuse for that combination of reasons; it needn't be clear in a normal case which reason is more important, and I think it generally isn't. If I'm right then what we judge in the normal case cannot decide what we should say about the case of Jane, where both factors are supposed to be efficacious.

I do not discount off-hand the fact that Jane really wanted to do what she did. Her wanting to do that was after all, we're assuming, efficacious in bringing about what she did. If this were all we knew, then what better reason could there be for holding her responsible? However, this isn't all we know, and what else we know—that her action was brought about because of causal determinism and this was sufficient to bring it about—makes this a case we don't easily understand, and don't, I think, have intuitions for.

There is also a reason to count Jane's admission that she really wanted to do what she did a little differently than we count the claim of the man who did what he did only because he could not have done otherwise, that he did not do what he did because he wanted to. This is that there is no reason that the efficacy of causal determinism in her actions would have made her aware that what she wants doesn't really make any difference to her actions, even if that were true. (We're assuming she's not a philosopher who's sat around becoming convinced of incompatibilism for years.) If deter-

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

minism is true, it's always been true, and even if it is incompatible with free will that doesn't change the fact that our practices of moral evaluation have evolved in utter disregard of that. Since there is no reason to think that our practices of evaluation and self-report would have already taken the incompatibility, if there is one, into account, we cannot simply assume that what Jane thinks about the efficacy of her will is correct.

This makes her case abnormal in another way, since I think that we do not normally doubt people's reports of what their wills or desires were, unless we think they are insincere or lying, but here we have another possible source of doubt about her testimony. Our question, in considering whether free will and determinism are incompatible, is not what we would say if a person testified to having done something of her own free will and we could assume the case is normal, or whether we could redescribe it as normal, but whether what she would testify and what we would say really mean what we think they mean, or indeed anything at all, if determinism holds, where we recognize that this scenario presents us with a view of our actions as abnormal.

Frankfurt's revised principle, that a person is not morally responsible for what he has done if he did it *only because* he could not have done otherwise, is true in the normal case, but does not decide the question for the abnormal case we must imagine when we take determinism seriously. The problem that arises in the attempt to view our actions both deterministically as events and as the consequences of our own wills is that the actions end up having more actually efficacious sufficient reasons than we are accustomed to cope with. Using Frankfurt's principle we could conclude that all is well with responsibility since we can still affirm that we did actions of our own accord (when we did) and not *only* because of causal determinism. But we could equally conclude that since causal determinism is sufficient, and actually efficacious, our impression that our wills have any efficacy in bringing about what we do 'of our own accord' is an illusion; since wills would be otiose, why should we believe they exist? If so, then we always do everything *only because* we are causally determined to do so, and Frankfurt's principle rather supports incompatibilism.

Strawson sees the question of whether determinism is compatible with free will, or responsibility, as the question whether determinism could imply that we should give up our reactive attitudes toward all people all of the time. He identifies, in the case of resentment for example, conditions under which we normally regard an injury as one for which the reactive attitudes are not appropriate and

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

asks whether determinism would imply that any of those conditions held for all injuries at all times. Is it the case that determinism would imply everyone is ignorant of causing injury every time they do, or had acceptably overriding reasons for acquiescing in doing so? Obviously not.[6] He applies the same test to the other set of conditions he has identified as ones under which we normally suspend reactive attitudes to a particular agent: would determinism imply that everyone suffered from deep-rooted psychological abnormality, that everyone is 'only a child,' that everyone 'wasn't himself lately,' that everyone has been 'acting under post-hypnotic suggestion' all of the time? Again the answer is that whatever the thesis of determinism is, surely it doesn't imply any of *that*.

Given what I have claimed about the problem juxtaposing free will and determinism presents us with, it will be no surprise that I take this to be an insufficient sort of defence of compatibilism. Because, as discussed above, the scenario considering our actions as both freely done and causally determined presents us with is precisely not normal, we do not address any problem it presents, or defuse any apparent problem automatically, by rehearsing examples of the sorts of conditions under which we normally suspend reactive attitudes towards people and asking whether determinism implies any of them. The level of detail Strawson resorts to here is a particularly weak aspect of his argument. No one ever thought determinism implied we were all insane, or all had just come back from the hypnotist. But the idea that there is a tension between free will and determinism is surely not as absurd as that. A more solid defence would go by way of identifying more generally why we take those conditions under which we excuse people from reactive attitudes to be excusing conditions, and arguing that determinism does not imply or resemble any of those reasons.

But this, I think, cannot be done. This is because for many of these excusing conditions, the reason they are excusing is either (roughly) that we think the person did what he did only because he could not have done otherwise or that we think the person did not do what he did of his own accord (or both). If so, then the question will become whether determinism implies either or both of those, about which I've already had something to say. Why do I excuse the insane man for pulling up my tulips, for screaming at the top of his lungs while I'm trying to work, for strangling my pet fish with his bare hands? Why do I suspend reactive attitudes (if I can)? It's because, and to the extent that, I view him as someone who does not

[6] The case of accident is interestingly omitted when it comes to applying this test, though it is mentioned when the conditions are first identified.

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

direct his own actions, because I believe he's a kind of thing which has something going on inside him he has no control over, a thing for which the notion of controlling itself may not even make sense, a thing which simply cannot behave otherwise than it does. 'No one is home' in him really, so there's no one to blame. He is a 'phenomenon', a set of events. This is a more plausible, and fuller, description of why I suspend reactive attitudes, and adopt objective attitudes, towards the insane man than simply that he is 'deeply abnormal.'

Strawson addresses these complaints of mine by claiming that there is no sense of 'determined' such that 'our adoption of the objective attitude towards the abnormal is the result of prior embracing of the belief that the behaviour, or the relevant stretch of behaviour, of the human being in question *is* determined in this sense.'[7] It may be true that it is not because of a belief in determinism, and a belief that the insane man's behaviour was determined in this sense, that we do, or would say we do, suspend the reactive attitudes. That would be claiming too much. However it's not enough to say, as Strawson does, that our reason for suspending reactive attitudes is that the man is 'incapacitated in some or all respects for ordinary interpersonal relationships.'[8] This is not enough because it simply repeats our claim that he is not up to being a target for reactive attitudes. We knew that, and we wanted to know what it was about him that made us regard him so. We are allowed, according to Strawson, to say further things about why our man is incapacitated—perhaps he lives in a fantasy world, or is an idiot, or is a moral idiot—but these things still do not tell us why we suspend reactive attitudes. Why is an idiot, or someone who lives in a fantasy world, excused from the praise and censure others must endure?

A fair description of why we suspend reactive attitudes in such a case is that when we ask the question 'Why did he do what he did?' or the question 'Why does he do what he does?' the only possible kind of answer is an objective one in which we see the person as a thing which does such and such. We may not be able to give any answer at all to this question in cases where the behaviour is unpredictable and/or the condition is not understood, but if we can, then the sort of answer we will give will point out that the man does not do what he does for reasons, he just does it. He is subject to something which makes him do thus and so. He can't help it. He simply does what he does and cannot do anything but what he, in fact, ends up doing. The things he does simply happen, that's all.

If Strawson's point were that our ordinary view of objects or

[7] Peter Strawson, 'Freedom and Resentment,' *Free Will*, Gary Watson (ed.) (Oxford University Press, 1982), 59–80, on 69.

[8] Strawson, 69.

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

things does not have determinism explicitly wrapped up in it, then we would have to agree. But we wouldn't have to be very impressed, because I do think our ordinary view of objects involves an idea that they lack the ability to do otherwise than they do. While it can be left to metaphysics to decide whether it is 'possible' for objects to do otherwise than they do, we do not believe that if it turned out that they 'could' do otherwise than they do, it would be because of any ability, so far neglectfully left unattributed to them, to *bring it about* that what happens should happen otherwise. Objects, as we view them, do not have the ability to do otherwise than they do, though it may be possible for them to do otherwise than they do. It is because, or to the extent that, when we ask of the insane man 'Why does he do what he does?' there is no sensible way to answer that question except by viewing him as an object in this sense, that we excuse him.

Notice that the reason I say we suspend the reactive attitudes toward the insane man has two parts, one part to do with his lack of agency, the other part to do with the fact that only one way of viewing him is left to us, as an object which among other things doesn't have the ability to do otherwise than it does. It may be that lack of agency implies that a person doesn't have the ability to do otherwise than he does, but in case it doesn't—think of a sane and intelligent dog—let's keep these notions separated. Notice that the way I've described the reason we suspend reactive attitudes toward the insane man does not make a claim that he could not have done otherwise any part of the reason for excusing him. However, the claim that a person could not have done otherwise implies that she doesn't have the ability to do otherwise, and since the first claim arguably follows from most forms of determinism, so does the second. Since the second claim is part of the reason we excuse a person *tout court* from being a target of reactive attitudes, or so I've argued, we see that determinism has some claim to produce the same effect.

This doesn't imply that determinism gives us reason to suspend all reactive attitudes toward all people all of the time, because I have not shown that it follows from determinism that the other condition I say is involved in our suspension of reactive attitudes holds, namely that the person lacked agency. However it does show, on grounds which Strawson ought to recognize as the right sort, that it is not absurd to think determinism would imply we should suspend all reactive attitudes.

About the question whether determinism implies or suggests that the other condition for suspending reactive attitudes holds, I refer to my point above that the case which the thesis of determinism requires us to entertain is one in which both my free will and causal determinism are to be both sufficient to bring my actions about and

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

actually efficacious in bringing them about. This scenario produces a crowding effect, in which if I believe one of those two factors is behaving as specified I cannot see why the other should be there at all, and I begin to doubt that it really is. Since one is in the process of assuming determinism to see what it implies or doesn't, causal determinism isn't the half that ends up looking dubious; the structure of the inquiry doesn't allow that. The only option left that will bring sense to the scenario is then to doubt that my free will is really efficacious. In this way the thesis that causal determinism is efficacious and sufficient to bring about my actions crowds out the efficacy of my agency. An agency that isn't efficacious isn't worth saying we possess, so the other condition required for us to suspend reactive attitudes begins to seem plausibly to obtain.

While Strawson is right that 'when we do in fact adopt [objective attitudes] in a particular case, our doing so is not the consequence of a theoretical conviction which might be expressed as "Determinism in this case.",'[9] I have argued that what we do is something importantly close to that. A belief that a person does not have the ability to do otherwise than she does is part of the reason that we suspend reactive attitudes in many cases—I cannot understand what 'she couldn't help it' means if it doesn't include that. The thesis of determinism implies that a person cannot do otherwise than she does, which in turn implies that she hasn't the ability to. Thus determinism would give us some reason to suspend reactive attitudes. The fact, as I've argued, that determinism crowds out the efficacy of our agency lends weight to the idea that determinism could mean the other condition for suspending reactive attitudes also obtains: that the person lacks agency. Thus determinism is a proper threat to our reactive attitudes.

Strawson offers another argument which is meant to supplement the first, an argument to the effect that suspending all reactive attitudes and treating everyone always and only by means of objective attitudes is not something human beings would be capable of, 'even if some general truth were a theoretical ground for it.' This is because the human commitment to interpersonal relationships is 'too thoroughgoing and deeply rooted' for us to be able to cease to participate in them, and being exposed to the reactive attitudes, as Strawson has described it, is essential to human interpersonal relationships. Because I worry that believing we are incapable in this way could be a result of lack of imagination, I would rather not rest too much on this and am interested in whether there could be a more theoretically grounded limit on our ability to believe the thesis of determinism.

There is reason to think that there is, having to do with the nature

[9] Strawson, 69.

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

of belief. I will argue that the *act* of believing that determinism is true involves something which would be excluded by the truth of determinism. Thus, I will argue that we can believe the thesis of determinism only if it's false. In many cases, believing something involves choosing among alternate possibilities. Did the criminal go into the bank or into the store? I believe she went into the store. Did she go into the bank or not? I believe not. Believing can involve a choice among alternate possibilities even when the alternatives are not stated: Did she go into the bank? I believe not. There may be beliefs for which this claim isn't true. I believe that the table in front of me has a blondish colour, but the alternate possibilities haven't as much as crossed my mind. I have simply believed what was actual, by inspection, and may not until this moment have even realized I had that belief. Whatever we might want to say about this second kind of belief, though, the thesis of determinism is not of this kind. Since it is a highly articulated thesis, one must have thought about it and at least one alternative to it, in order to believe it. And when one is believing it or not, one is choosing between those alternate possibilities.

Or is one? Let's suppose without argument that the thesis of causal determinism implies that for everything that happens it could not have happened otherwise than it does. This implies, I take it, that there are no alternate possibilities, anywhere, in anything that happens. In particular there are no alternate possibilities called 'determinism' and its alternative. Since one cannot choose between alternatives which don't exist, and believing, in a case like believing the thesis of determinism, requires choosing between alternatives, it follows from the thesis of determinism that believing the thesis of determinism never happens. It can happen that people say the words 'I believe the thesis of determinism.' but merely saying words, we've assumed, is not enough for believing to be going on, at least in such a case. Since the thesis of determinism cannot be believed unless it's false, it follows that we cannot believe that thesis unless it's false.

If this argument is successful, then Strawson is right to think we can't believe the thesis of determinism even if it's true, not because of its consequences, but because what is required for there to be believing of theses like determinism is not allowed by the thesis of determinism. We should rather say that if determinism is true, we cannot believe it *because* it's true. What if the thesis is false? In that case we could believe it, but we would be mistaken to do so. And if we reflected a moment we would see that if we really are believing it, then it must be false. But admitting that a belief is false is tantamount to ceasing to believe it. That we can believe the thesis of determinism obviously does not imply that we should. If we can believe the thesis of determinism then that thesis is false, so if we

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press

think we can believe that thesis—whether we actually do believe it or not—then we think that thesis is false. It seems that the only possible belief we can have about the matter of determinism which does not undermine itself is that that thesis is false. If we were to believe it is true, then we would be believing something which implies that 'this is not a belief'.

I am not sure whether we can believe the thesis of determinism, for two reasons. First, I'm not sure that what we would be doing when we affirmed it would be believing in the sense I've specified, because I'm not sure there *are* alternate possibilities.[10] But secondly, and from a different perspective, I'm not sure that we have ever been successful in entertaining what the scenario of determinism is when applied to ourselves and combined with things we normally assume about ourselves. This is related to the point I made several times earlier that when we try to imagine what it means to say that both our free wills and causal determinism are both efficacious in bringing our actions about and sufficient to bring them about, this doesn't make sense, is overcrowded, and inclines us to discount the one or the other depending on our interest on a given occasion. This suggests that we don't quite know what it would be to imagine universal determinism, and if so then we are probably not believing it either, at least not yet, even when we think we are. All of this suggests that we cannot, even on the strength of the argument I just described, simply verify that determinism is false by verifying that we can believe it, for the latter verification is not a patently obvious affair.

Strawson's second argument essentially concludes that even if determinism *were* incompatible with, or undermining of, our reactive attitudes (and were true), we would not be able to believe it. My argument concludes that if determinism is true then we cannot believe it. These arguments do not prove incompatibilism, but consistently with everything I have stressed above, they do not confirm compatibilism either. If anything, my final argument is suggestive of incompatibilism, since the reason that we cannot believe determinism if it is true is that what is involved in the thesis of determinism being true disallows something we normally take for granted about ourselves, namely that we have the ability to believe theses like determinism. Thus it shows the capacity of the thesis of determinism to have powerfully revisionist implications.

Harvard University

[10] Note that my saying this is not necessarily self-undermining (though it might actually be). I can affirm or deny what I like as long as doing so does not require the existence of alternate possibilities. Maybe I only think it does, or maybe what I've just entertained happens not to.

https://doi.org/10.1017/S0031819198004033 Published online by Cambridge University Press