Strawsonian Optimism for Libertarians

Abstract

In this paper, I defend the idea that libertarians may be 'Strawsonian Optimists' – that is to say, that they may consistently hold that even if determinism is true, there need be no threat of any kind to the concepts and practices constituting our commitments to moral responsibility and personhood. My defence of this position takes place in the context of John Fischer's attempt to defend what I call the 'Invulnerability Intuition' – that is to say, the intuition that it would not be the case that we ought to give up our commitments to moral responsibility and personhood if we were to be told unequivocally one day by theoretical physicists that there is simply no doubt that the universe is deterministic. Fischer not only defends the Invulnerability Intuition; he also insists that the libertarian cannot accept it – and criticizes what he takes to be van Inwagen's attempt to claim otherwise, characterizing van Inwagen's strategy as a variety of 'metaphysical flipflopping'. In this paper, I not only insist that flipflopping can be acceptable – I claim also (i) that flipflopping is easier to defend if one is an Agency Incompatibilist, than if one is a more standard kind of libertarian; and (ii) that flipflopping is a preferable strategy for saving the Invulnerability Intuition than Fischer's own strategy – the metaphysics of semicompatibilism.

John Fischer has brought an immense amount to the philosophy of free will and moral responsibility. There is no modern philosopher working in this area from whom I have learned more and whose arguments have always seemed so thoroughly worth engaging with, even where I have disagreed vehemently with their conclusions. One of John's most important virtues, I think, is that he has an uncanny knack for crystallising deeply intuitive points which are (or ought to be) at work in certain important dialectical contexts, but which have somehow gone unnoticed or unexpressed by others. A particularly notable example of the genre of point-crystallisation I have in mind is John's isolation of the idea of a mere 'flicker' of freedom, an alternative possibility which however could not be the sort of thing that might ground the attribution of moral responsibility to a person allegedly in possession of it, and hence which (in John's view) could not play the role that some hoped it could play in the debate. This intervention served to bring much-needed clarity to a discussion where it was badly needed, clarity which went on to transform the libertarian landscape.

In this paper, I want to take a look at another deeply intuitive point which has played an important role in Fischer's thinking concerning free will and moral responsibility, and the relation between them. In some ways, indeed, it has been the guiding intuition behind the overall shape of the view which he has come ultimately to

endorse; and it is also an intuition whose attractions I very much recognise and understand. For the purposes of this paper, I am going to call it the 'Invulnerability Intuition' and it concerns the question how we should think about a certain imaginary future situation in which the truth of causal determinism has been scientifically proven beyond reasonable doubt.

The imaginary future situation in question is described by Fischer here:

> Suppose … that a consortium of well-respected scientists announce that they have developed a remarkable new theory which implies that all events can in principle be fully explained by previous events and the laws of nature. That is, they claim that although they cannot at present make all the predictions about the future, their theory implies that the world is *not* fundamentally indeterministic as many scientists had previously thought; rather, if one knows enough about the past states of the world and the laws of nature, one can confidently predict all the states of the world in the future (1994: p.6).

The Invulnerability Intuition insists that in the scenario described (which I shall take the liberty of calling – somewhat provocatively - the 'Nightmare Scenario'), we would not be inclined to give up, and moreover it is not the case that we *ought* to give up[1] any of the practices which together constitute our commitment to what Fischer calls 'the distinction between persons and non-persons'. According to Fischer, these commitments include:

(i) the idea that persons have a particularly stringent right to existence as compared with non-persons;
(ii) the claim that it is sometimes appropriate to harbour certain attitudes towards other persons (in particular, the so-called 'reactive' attitudes, examples of which might be resentment, gratitude and forgiveness, for instance), in virtue of the attitudes and intentions they display towards you;[2]
(iii) the idea that a range of social and institutional practices, including praising, blaming, punishing, forgiving and rewarding one another which are, on Fischer's view intimately bound up with the reactive attitudes, are, in general, appropriately well-justified.

It is the normative claim, the claim that it would not be the case that we ought to give up this distinction between persons and non-persons in the Nightmare Scenario, rather than the merely psychological claim that as human beings we

[1] There is a difference, of course, between the claim that 'in such and such circumstances, it would not be the case that we ought to give up X' and the claim that 'in such and such circumstances, it would be the case that we ought not to give up X'. It may be that this second, stronger version of the Invulnerability Intuition is defensible and I believe P.F. Strawson certainly meant to defend it in his (1962). But I am much less sure that Fischer means to do so; and I therefore stick here with the weaker version.
[2] See Strawson (1962) for the elaboration of the concept of 'reactive attitudes'.

are in fact *incapable* of dispensing with the distinction, that will interest me here. Fischer puts the point as follows:

> I am saying that, upon due reflection, it just does not seem appropriate or plausible to think that we should abandon our view of ourselves as persons, if it turned out that the consortium of scientists were correct.
>
> Although I am not sure how precisely to articulate the basis of this point, it does seem to be strong. And … I am not here claiming that the normative point is obviously correct; I am merely pointing out that it has a rather strong intuitive basis. Our relationships with our family and friends are extraordinarily significant to us … It is very hard to see how the discovery of the consortium of scientists should move us to give up these features of our lives" (1994, p.7).

In this paper, after locating Fischer himself within the tradition which defends the Invulnerability Intuition, my aim will be to explore other – and in particular, libertarian - possibilities for its defence. I agree with Fischer when he says that he is unsure what the intuitive basis of the Invulnerability Intuition might be – and I do not want to go so far as to argue that the intuition is definitely correct. What I am concerned to show instead is that libertarians need not necessarily be committed merely by their libertarianism to the claim that it is definitely *in*correct. For – as Fischer himself has argued[3] and as I shall shortly explain - it can be difficult to see how any libertarian about moral responsibility could avoid rejecting the Invulnerability Intuition without adopting positions which might seem deeply implausible. I shall be attempting to offer here a new kind of libertarian response to the difficulty.

The Invulnerability Intuition has unmistakeable roots in the work of P.F. Strawson, and the defence of at least one version of it can be regarded as one of the main aims of his seminal paper, 'Freedom and Resentment' (Strawson, 1962). Fischer's version of this intuition, however, is importantly distinct from what it seems to me Strawson intended to recommend. Strawson distinguished in his paper between the 'pessimists' in the free will/moral responsibility debate who hold that if the thesis of determinism is true, the concepts of moral obligation and responsibility lack application, and the associated practices of punishing, blaming, etc. are really unjustified; and the 'optimist', who holds that these concepts and practices in no way lose their justification, if determinism is true. Strawson noted the widely felt insufficiency of the (then) very common 'optimist's' suggestion that we might seek to base the justification of punishment, blame, etc. merely on their overall utility, conduciveness to good behaviour, orderly societies, and the like; and suggested that his invocation of the reactive attitudes might 'give the optimist something more to say'. It is an important question, though, which I will consider below, whether the use

[3] Fischer's first critique of what he calls the 'metaphysical flip-flopping' he takes (rightly) to be essential to any libertarian attempt to secure what he calls 'resilience' is to be found, briefly considered, in Fischer and Ravizza (1998), 253-4. More developed considerations of what, precisely, might be wrong with the strategy are to be found in Fischer (2016) and especially Fischer (2024).

Strawson makes of the reactive attitudes can really provide the optimist with *enough* to say in the face of the insistence that, even granted that it would be (a) psychologically immensely difficult and (b) highly inadvisable in terms of its effects on human well-being, to dispense with commitments (i)-(iii) above, it would remain, in one very clear sense, irrational to continue to endorse those commitments, given the Nightmare Scenario. Fischer should be regarded, then, I think, as a philosopher who is attempting to respond to the imperative to say more than Strawson does about this particular concern. In that sense, we can regard him as someone who has given the optimist yet more to say than Strawson did, building on the Strawsonian basis of the reactive attitudes, but supplying an important story, absent from the Strawsonian picture, about how optimism in the face of determinism might be defended even in the face of potentially very strong arguments to the effect that our access to alternative possibilities, and hence our possession of free will, is inconsistent with it.

It is natural to think that Strawson's 'pessimist' must be identical with the incompatibilist – and indeed, I am sure that is what Strawson intended to imply. The incompatibilist who also believes in free will and moral responsibility, it is true, is unlikely to think of themselves as any kind of pessimist about those things – but what Strawson means by 'pessimism' in this context seems to be pessimism about *compatibilism* rather than pessimism about moral responsibility. And one might think that of course the incompatibilist is bound to be a pessimist about *that*. Nevertheless, my aim here is to argue that strictly speaking, an incompatibilist need not actually be a Strawsonian pessimist about compatibilism, defined as Strawson defines it, at all. My claim will be that a certain kind of perfectly defensible optimism remains available to the incompatibilist who wishes to leave the door ajar for the Invulnerability Intuition. Moreover, I shall argue that this kind of optimism is actually able to support a more robust and intuitive account of the distinction between persons and non-persons than Fischer himself can be confident of being able to offer.

One might wonder, though, as Fischer himself wonders, how any libertarian could possibly turn this trick. In my own particular case, the situation is this: I have maintained both that having (what I call) 'agency' is a necessary condition for possessing moral responsibility, and also that agency could exist only in an indeterministic world, because it is itself an essentially indeterministic phenomenon. This latter claim is at the heart of the view I call 'Agency Incompatibilism', which I develop and defend in my (2012). (Many other libertarians would, I think, agree with these two claims, provided the word 'free', which I regard as otiose in this context, were added before 'agency'). But if these two claims are correct, they seem to imply that moral responsibility could *also* exist only in an indeterministic world. Hence, it seems logical to suppose that I (and most other libertarians) must be committed to the view that, were the Nightmare Scenario to come to pass, the collection of important ideas and practices that Fischer collects under the head of 'moral responsibility', would have to be given up (where the 'have to' here is intended to be that of rational necessity, rather than of merely psychological compulsion). The only

possible alternative might seem to be to deny that the Nightmare Scenario *could* actually come to pass – an option which, for reasons I shall shortly explain, I do not think it would be right to endorse. This paper, then, is my attempt to explain why, despite initial appearances, I think there is a better option than this available for the libertarian. I made a start on this task in my (2012)[4]; but there is more to say.

The basis of the strategy I shall defend has already been the subject of some discussion in the literature. The best-known version of the strategy is van Inwagen's. Van Inwagen writes as follows about his preferred response to the coming about of the Nightmare Scenario:

> … it is conceivable that science will one day present us with compelling reasons for believing in determinism. Then, and only then, I think, should we become compatibilists, for, in the case imagined, science has *ex hypothesi* shown that something I have argued for is false". (van Inwagen, 1983, p. 223)

Van Inwagen's suggestion, then, is that should the Nightmare Scenario come to pass, one should become a compatibilist *at that point.* Fischer calls this strategy 'metaphysical flip-flopping' and has made a number of important objections to it. However, a very good beginning to a defence of flip-flopping against some of these objections has already been provided by Bailey and Seymour (2021), some of whose arguments I endorse. Fischer has, however, replied to Bailey and Seymour in his (2024), and one of my main aims here, therefore, will be to respond to these latest objections. It is important to note, though, that I shall be offering a response on behalf not of van Inwagen (who is usually Fischer's main target) but rather on behalf of the brand of libertarianism I myself favour, Agency Incompatibilism.[5] It will matter to some of the arguments I shall make later that the overall shape of, and case for, Agency Incompatibilism are importantly different in several respects from the shape of, and case for, van Inwagen's version of libertarianism, and indeed, it is very different from *most* varieties of libertarianism in significant ways. These differences, I shall suggest, can enable Agency Incompatibilism to offer more effective rejoinders to some of Fischer's arguments against the dialectical acceptability of 'flip-flopping'.

In section (i), I shall try to say a bit more about what the correct interpretation of Fischer's version of the Invulnerability Intuition should be and will remind the reader of how Fischer himself attempts to secure the necessary invulnerability. In section (ii), I shall begin my consideration of the question whether someone who accepts both Agency Incompatibilism and the claim that agency is a necessary condition of moral responsibility, could manage nevertheless to avoid contradicting the Invulnerability Intuition. I consider, in particular, the possibility of denying the conceivability of the Nightmare Scenario – but despite its attractions, I shall in the end conclude that that is too hard a bullet to bite. In the third section, therefore, I

[4] See in particular Chapter 5, 'The Epistemological Argument'.

[5] I set out the fullest case for this view in my (2012). Other papers relevant to the defence of Agency Incompatibilism are my (2008), (2009), (2011) and (2016).

shall suggest that the response that Fischer calls ‘flip-flopping’ and which Bailey and Seymour characterise rather more favourably as ‘responding to new evidence’, is indeed the way to go for the libertarian. In section (iv), I’ll then defend this solution against Fischer’s objections, in the context especially of Agency Incompatibilism, before suggesting in section (v), by means of some admittedly imperfect but nevertheless suggestive analogies, that as compared with my libertarian solution, Fischer’s own strategy makes unacceptable concessions for the sake of rendering the Invulnerability Intuition secure.

(i) *Fischer on the Invulnerability Intuition*

The Invulnerability Intuition is the claim, recall, that were the Nightmare Scenario to come about, it is not the case that we *ought* then to give up the collection of beliefs and practices which Fischer collects under the concept of ‘moral responsibility’. It is clear, then, as I have already said, and as Fischer tells us explicitly, that Fischer’s claim is intended to be *normative* and not merely psychological. But there is still disambiguation to be done, for there remains an important question about what sort of normativity is involved here. Does Fischer intend to claim, for example, that, supposing for the sake of argument that it would be psychologically possible to do so, it would nevertheless still not be *rational* to give up the collection of beliefs and practices in question? – and if so, what kind of irrationality would be in question?

The question what kind of rationality would be at stake here in any claim to the effect that it would be irrational to maintain the beliefs and practices relating to moral responsibility if determinism were true is addressed explicitly by Strawson in ‘Freedom and Resentment’. Strawson makes clear, I think, that he is of the view that the rationality at issue here could not be the purely epistemic variety which relates theoretical beliefs to the equally theoretical reasons for them:

> … if we could imagine what we cannot have, viz., a choice in this matter, then we could choose rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth or falsity of a general thesis of determinism would not bear on the rationality of *this* choice (1962, p.83).

At a later point in the same paper, he further comments, with respect to the closely related suggestion that we might give up what he calls the ‘vicarious analogues’ of the interpersonal attitudes (things like moral indignation which can be felt on behalf of another, not only on behalf of oneself) in the face, say, of a proof of determinism, that “if there were, say, for a moment open to us the possibility of such a godlike choice, the rationality of making or refusing it would be determined by quite other considerations than the truth or falsity of the general theoretical doctrine in question” (1962, p.87). What Strawson appears to be saying is that the rationality which is in

question, when we consider the issue whether the truth of determinism could make it rational for us to discard the interpersonal and analogous vicarious attitudes, *could only be* the kind of rationality which accepts as reasons claims about such things as the enrichment or impoverishment of human life – a kind of rationality that I would be inclined, for present purposes, to call 'practical' (in implicit opposition to 'theoretical') though the label is perhaps not altogether apt.[6] The important point for my purposes, though, is that Strawson is insistent that there is simply no genuine place here for the relevance of any distinctive *theoretical* rationality which might, for example, impugn the idea that the practices of punishment, blame, etc. could be *fair*, given the truth of determinism and its possible implication that no one is ever able to do other than they do.

It is less clear, though, why exactly Strawson believes this. Might not someone convinced of incompatibilism about determinism and moral responsibility, and also about the truth of determinism, legitimately take the view that even though it might be both psychologically impossible and 'practically' irrational (in the sense that it would make human life immeasurably worse) to give up commitments (i)-(iii), there would still be a sense in which it remained *theoretically* irrational to continue to hold the relevant beliefs – that they would then be lacking in theoretical justification? Strawson's answer is obviously 'no' – but a clear argument is wanting. Considering this very question, he says only this:

> "… such a question could seem real only to one who had utterly failed to grasp the purport of the preceding answer, the fact of our natural human commitment to ordinary interpersonal attitudes. This commitment is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within the general framework." (p.83)

But one might wonder why we are not allowed to question whether the general framework of human life might be based in certain ways on beliefs which, once spelled out, might appear to us, for various reasons, to be *false*, in virtue of certain things that we have come to know. Could we not do so even while accepting, with Strawson, that that general framework is almost certainly here to stay – and moreover, that that is very probably a good thing?

---

[6] The reason for thinking it may not be entirely apt is that the realm of practical rationality normally takes for granted some conception of the 'ends' of activity and then considers the best means of achieving them. Whereas what is at issue here relates to the question whether it might be rational to discard a certain profoundly entrenched set of beliefs, attitudes and practices which structure the forms taken by human social life. That is a question intuitively deeper than the merely practical – although it certainly relates, in this context, to a question about what we should *do*. Cf Frankfurt's disentanglement, in the first chapter of his (1982) of the question 'what to care about', from the domain not only of epistemology (the realm of 'theoretical' rationality and justification) but also from ethics (the realm of 'practical' rationality and justification), whose central question is 'how to behave'. (Frankfurt 1982, p,80). Strawson seems to be making a point about what to care about.

For one who feels that they have been left at the end of 'Freedom and Resentment' with a question that has not been answered by Strawson's trenchant defence of our commitments to ordinary interpersonal attitudes, it is pertinent to point out that Fischer's interpretation of the Invulnerability Intuition seems different from Strawson's. Recall that in insisting that he intends to make a normative claim, Fischer goes on to clarify what he is saying in the following way: that "it just does not seem *appropriate* or *plausible* (my italics) to abandon our view of ourselves as persons" (1994, p.7). 'Appropriate' is admittedly an all-purpose normative workhorse – but in 'plausibility' we have, I think, an unmistakeable connection to *theoretical rationality*. What is plausible is surely what is plausibly *true*. (I am also encouraged by the word 'view'; it is not (or not only) practices or attitudes, but a 'view' of ourselves that seems to be in question, for Fischer). Whatever else Fischer might mean, then, I think it is pretty clear that he is saying, in a way that Strawson does not straightforwardly say, that even if the Nightmare Scenario were to come to pass, we would still not be *theoretically justified* in giving up the crucial beliefs that constitute the heart of the notion of moral responsibility (for example, that people sometimes deserve to be punished and that I am sometimes well-justified in blaming another person for what they have done to me or to someone else). Fischer's version of the Invulnerability Intuition, then, seems different from Strawson's own. There is a sense in which, for Strawson, the incompatibilist challenge to moral responsibility is no challenge at all, once we come to see the situation aright. For Fischer, though, the Invulnerability Intuition must meet head-on the challenge offered by incompatibilist arguments such as van Inwagen's (1983) Consequence Argument, which purport to show that if determinism is true, there is no free will (because if determinism is true, there are no alternative possibilities of the kind that would be required for it). What if there was no free will? What if nothing anyone does is ever up to them? How can the Invulnerability Intuition be maintained in the face of such possible theoretical discoveries? For Fischer, the Invulnerability Intuition is required to run the gauntlet of this range of questions and yet survive.

Fischer's solution to the issues posed by this set of questions, as is well-known, is to argue that even if van Inwagen and others may be right about the incompatibility of *free will* and determinism (a question about which he wishes to remain officially agnostic), we can and should retain our right to endorse the Invulnerability Intuition by giving up the traditional compatibilist acceptance of the idea that free will and moral responsibility stand or fall together – and hence that if moral responsibility is to be compatible with determinism, free will must be shown to be so, too. The result is Fischer's distinctive metaphysics of semi-compatibilism – a strategy designed to *insulate* the Invulnerability Intuition from attack from incompatibilist arguments about free will. Given semi-compatibilism, it can be argued that moral responsibility is simply invulnerable to anything that could be thrown at it by the Nightmare Scenario, because it depends only on a variety of agential control, guidance control, which does not demand any full-blown alternative possibilities, and is thus perfectly compatible with determinism. Fischer explains his distinction between regulative and

guidance control by reference to the example of driving. Suppose I am driving a car, and it is functioning well. I want to turn right – and therefore I turn the wheel to the right. The car does what I want in response to my action; I guide it to the right, and thereby have what Fischer calls 'guidance control' – the car does what I intended it should do in response to my action. We may still be uncertain what would have happened, however, in the event that I had wanted to turn left. Did I also have the ability to guide the car to the left? Perhaps – if the car was functioning properly. But suppose it had not been functioning properly because some essential part of the steering mechanism was broken – some part essential to the capacity of the car to be moved left by the steering mechanism? In that case, I would not have been able also to guide the car to the left – and therefore did not (though perhaps unbeknownst to me) have the power to move the car in whichever direction I wished. This, though, need not affect the truth of the claim that I in fact guided the car to the right, according to Fischer. In his terminology, I had guidance control, even though I did not have regulative control. Fischer concedes that regulative control may be essential for *free will* and hence that it may be the case that free will is incompatible with determinism because perhaps it requires real, alternative possibilities of the 'forking paths' variety that Fischer has done so much to help characterise. But regulative control is not required *for moral responsibility*, according to Fischer. All we need to know in order to justify the reactive attitudes, punishment, praise and blame, etc., is that the agent was appropriately reasons-responsive in whatever case is before us – and reasons-responsiveness of the relevant sort does not require the future to be genuinely open.[7]

Unlike Strawson, though, Fischer does not attempt to claim that the truth of the thesis of determinism would not bear *in any important way* on our views about persons. On the contrary, he explicitly allows that sufficiently strong arguments might perhaps show, for example, that persons do not possess the power of regulative control if determinism is true; and that therefore they might not possess free will under such circumstances. It is simply that our moral responsibility and personhood would not thereby be impugned. But for someone who wonders (as I do) why a being with no access to genuinely forking paths, a being whose so-called 'actions' were simply events produced deterministically by way of the conjoint influence of the laws of nature and prior conditions, would count as an agent and hence as a 'person' at all, Fischer's way of safeguarding the Invulnerability Intuition will seem to have failed in its overall task. This is a point, note, which gives the Agency Incompatibilist an argument *against* Fischer's semi-compatibilist strategy which is unavailable to most other libertarians. Most libertarians tend to accept that *some* actions may perfectly well be determined, separating off a special class of actions which are regarded as

[7] For those wanting further details of semi-compatibilism, Fischer's overall picture, including a detailed account of what 'reasons-responsiveness' consists in, is spelled out in a wide variety of books and papers, spanning many years, some co-written with others. There is an excellent summary of Fischer's overall view in Fischer, Kane, Vargas and Pereboom (2007). More detailed presentations can be found in Fischer (1994), Fischer and Ravizza (1998) and the essays contained in Fischer (2006).

'free' to be the topic of especial concern. Any libertarian who holds this view is blocked thereby from the utilisation of any argument which depends on the premise that a world in which determinism holds should be regarded as a world in which there are no agents. But this premise represents the heart of Agency Incompatibilism. For the Agency Incompatibilist, agency is *ipso facto* free,[8] as it were, because the phenomenon of agency is characterised essentially by involving the settling of at least some hitherto unsettled matters. Anyone holding this view will be committed to the claim that persons would not exist in a deterministic universe, because there would be no *agents* in such a universe (just as water would not exist in a universe in which there was no H_2O). The project of rescuing personhood by way of a strategy premised on ensuring that *all* agency might perfectly well be determined, for all we know, will seem, then, to be doomed to failure from the start, from this point of view. For the Agency Incompatibilist, then, incompatibilism is not just a matter, as it is for most libertarians, of providing a metaphysics which can sustain free will and moral responsibility. It is a matter of providing *for the metaphysics required by personhood itself* (at least on the very plausible assumption that persons are necessarily agents). That makes semi-compatibilism, with its attempt to save personhood *independently* of saving free will, a complete non-starter.

(ii) Can an Agency Incompatibilist save the Invulnerability Intuition?

The Agency Incompatibilist maintains that agency itself, conceived of simply as the power to *act*, is incompatible with determinism: that nothing would be capable of action which had no access to alternative possibilities of the robust sort which determinism rules out. Having the power to act, moreover, is plausibly a necessary condition of having any moral responsibility for anything. (I will not defend this second assumption, since I think it would be common both to Fischer and myself). It follows, however, from these two claims, that moral responsibility is incompatible with determinism. Since Fischer would deny the first premise of this little argument, on the grounds that agency does not require alternative possibilities of the relevant robust kind, he need not worry about this conclusion. But *I* need to worry about it, if I want to hang on to the Invulnerability Intuition. How can an Agency Incompatibilist avoid contradicting this intuition? How is she to insist that *even if* scientists were to give us incontrovertible scientific evidence that determinism is true, it would *still* not be appropriate or plausible to conclude that no one was morally responsible for anything?

The natural thing for the Agency Incompatibilist to fall back on is the claim that the antecedent of the conditional in question will never be satisfied – that is, to insist that

[8] Though this is not a matter of mere *definition*. We are talking, as it were, about real and not nominal essence – and this is important. I discuss this further below.

the Nightmare Scenario will never materialise. An implicit assumption of the debate as generally characterised, she may point out, is that determinism may be true *for all we know*. But she will insist that this is a false assumption. For we certainly know, she may say, that there are agents; and we can also come to know, by means of philosophical reflection, that agency is incompatible with determinism.[9] It is part and parcel, she will insist, of our conception of agency, that agents are settlers of matters – that they are entities which can make it the case that things go a certain way in the world when those things *could* have gone a different way. We know, she will say, even if we have not made the knowledge explicit to ourselves, that this world is a world in which things are settled *in time* (by us), not merely settled in the beginning, *for all time*.[10] The Agency Incompatibilist may therefore insist that we may use the known fact of agency, and its (she believes, knowable) incompatibility with determinism, to argue for the claim that we know that the Nightmare Scenario could never possibly come to pass, because we already implicitly know that the truth of determinism is incompatible with *other* things that we may justifiably claim to know.

However, it would be a bold philosopher who would unhesitatingly endorse such an argument. Science has provided evidence over the centuries for so very many counter-intuitive and virtually inconceivable things. Even if I am right that that the falsity of determinism ought properly to be regarded as part of our foundational world view, it might be pointed out that our foundational world view has been shaken over and over again – for example by the General Theory of Relativity, and by some of the claims made in Quantum Mechanics. It surely cannot be ruled out *a priori* – even if it is highly unlikely – that science will show us one day that there is no alternative to a deterministic picture. It seems simply wrong to claim that this scenario is inconceivable. Regretfully, therefore, even though I *do* think that we are already in a position to know that determinism is false, I do not think that it is completely inconceivable that it is true. Note that there is nothing paradoxical about this position. Merely being in a position to know things does not imply the *inconceivability* of their turning out to be false. All but the sceptic must admit, indeed, that *most* things we can legitimately claim to know are like this. I believe I know who my parents are, for example, but I could conceivably be wrong. It might turn out that I have been the victim of a lie or deception. To take a more metaphysical example, I know that the external world exists, but many philosophers would accept that this is

---

[9] Timothy O'Connor (2019: 106) has suggested that perhaps those who claim to disbelieve such theoretical propositions as this might in fact believe them (as shown, for example, by immersion in practices of many kinds which can be argued to presuppose them), while merely *believing* that they disbelieve them. I can't go into this here, but I rather like this suggestion.

[10] I do not here address the question of *how* we can know this. But my view is that it is a claim similar in its foundational nature to the claim that there is an external world. That there is an external world is a proposition that most philosophers believe we can claim to know even if we are not able definitively to rule out the various counter-possibilities (dreaming, evil deceivers, brains in vats, and the like). In my view, the claim that agency involves settling is an equally fundamental piece of knowledge, which is likewise not based on evidence but on its utterly foundational role in the very idea of a subject of mental states of the kind which imply activity, such as thinking, deciding, choosing, etc.

compatible with its being conceivable that I am wrong about this; perhaps I am a brain in a vat, or being deceived by an evil deceiver. And likewise, I cannot accept that the Nightmare Scenario is inconceivable, even though I claim to know things which would make it impossible for it ever to come about. Moreover, the Nightmare Scenario is only (epistemically) impossible *tout court* on the assumption that I am right in my arguments for incompatibilism and for the existence of agency. But it is far from impossible, of course, that I am not right. A certain kind of humility is appropriate in philosophy in general – moreover, it is especially appropriate if one goes out on a limb and defends a position which is very unusual or at odds with what many others maintain. As well as the first-order case for Agency Incompatibilism which I have made in my work, then, I must take into account the meta-case that exists for supposing that this first-order case may well be flawed, i.e. the existence of thousands of intelligent compatibilists who believe that agency is perfectly compatible with determinism![11] And since such a meta-case can be constructed, it seems at least conceivable that the Nightmare Scenario might arise. It is true that for the Agency Incompatibilist, it is not correct to say that determinism is true *for all we know*, since she thinks we *do* (or should) know things which contradict it. But it would not follow from this that the Nightmare Scenario is inconceivable. And if it is not inconceivable, its inconceivability does not offer the Agency Incompatibilist a way to defend the Invulnerability Intuition.

(iii) *The 'Flip-flopping' Solution*

If the Nightmare Scenario is not inconceivable, then, what should the Agency Incompatibilist say, if it were to come to pass? What should she say it if were incontrovertibly shown scientifically, beyond reasonable doubt, that determinism was true? One possibility, of course, is that she might simply say that since agency had now been shown not to exist after all, that Fischer's commitments (and her own) to such things as the distinction between persons and non-persons and the appropriateness of the reactive attitudes would all have to be given up. Another would be for her to insist doggedly that the scientists must have got it wrong, compelling scientific evidence notwithstanding. But neither of these options seems at all attractive to me. I want therefore to explore and defend the availability of a third possibility – the possibility that the Agency Incompatibilist might instead simply acknowledge that this new scientific evidence had been an epistemic game-changer and that she has now been brought, by the new discovery, to accept the truth of compatibilism. As noted above, this move is not new – it is suggested by Van Inwagen, considering the very situation we have here been calling 'The Nightmare Scenario'. The overall idea is to offer a 'two-pronged' response to defend the

[11] Though it is highly questionable, in my view, whether some of these compatibilists really conceive of determinism in such a way that it ends up being a thesis that is even *prima facie* worrying for free will, because of their (Lewisian) conception of what a law of nature is. See my (2021).

Invulnerability Intuition. For the Agency Incompatibilist, the first 'prong' consists of arguing that we can already know by means of argument and reflection that determinism is false, because the existence of agency disproves it. The argument for this conclusion would have the following basic structure:

(P1) There is agency.

(P2) If determinism were true, there would be no agency.

So, (C) It is not the case that determinism is true.

This first stage constitutes the Agency Incompatibilist's continued endorsement of a libertarian position, and the case for P1 and P2, together with her acceptance of the argument above, constitute what she takes to be excellent reasons for thinking that the Nightmare Scenario will never actually arise. Since she takes the level of rational credence supplied to this conclusion by this argument to be extremely high, note, this is already sufficient to impugn Fischer's claim that moral responsibility 'hangs by a thread' on the libertarian view – it does not hang by a thread, in the view of the Agency Incompatibilist, but is rather supported by a huge steel girder. This is another respect, indeed, in which I regard the Agency Incompatibilist as better off, with respect to her capacity to respond to some of Fischer's concerns about the libertarian's defence of the Invulnerability Intuition, than many other kinds of libertarian. For it is important to note that Agency Incompatibilism takes a strong position on the justifiability of a negative *philosophical* verdict on the question of determinism *itself*. The Agency Incompatibilist takes it to be extraordinarily unlikely that determinism is true (at least partly because of the existence of the phenomenon of agency and her conception of its place in nature). She does not therefore accept the view taken by Fischer and endorsed by most other commentators on the free will problem (including many libertarians) that the question of determinism/indeterminism is simply one for the physicists to decide, a question which might easily be answered either way, in which case all philosophers can do is make judgements about the compatibility of each with free will and wait hopefully for the verdict of science. As argued in my (2012, chapter 5), it is essential to distinguish between the following two claims:

(D1) The question whether determinism is true is a question that can only be answered by physics.

(D2) The question whether determinism is true is a question that may (one day) be settled by physics.

I accept (D2). But I do not accept (D1). (D1) depends, in my view, on exceedingly strong claims about the bottom-up determination of reality which many phenomena, including the very phenomenon of agency itself, throws into question – and which may therefore perfectly well receive a well-supported negative verdict *from philosophers.*

So much for the first prong. The second 'prong' of the strategy consists of acknowledging that this argument nevertheless does not rule out the *conceivability* of the Nightmare Scenario[12], and of accepting that if it were to turn out that determinism is true, one ought (as a matter of theoretical rationality) to become a compatibilist about agency and determinism *at that point*. The second part of the strategy, then, constitutes the Agency Incompatibilist's considered decision that in such a situation, there would then be *stronger* rational grounds for maintaining the claim that agency (and free will) exist (one component of her original commitment to libertarianism) over the claim that agency is incompatible with determinism (the other component). Given the two-pronged strategy, there is therefore *no* situation in which the Agency Incompatibilist would be forced to give up her commitment to the existence of agency as a result of scientific discovery. Her conviction that agency exists would trump her conviction that incompatibilism is true, if push ever came to shove – not just psychologically, note, because of wishful thinking – but *rationally* – in the sense that she believes there would in such a situation be *stronger reasons* to maintain belief in agency, than to maintain belief in incompatibilism. To defend that is not quite to defend the Invulnerability Intuition, because the Invulnerability Intuition as we have it in Fischer relates to *moral responsibility*, not to agency. But it *is* to block the argument to the conclusion that the Agency Incompatibilist is committed to *denying* the Invulnerability Intuition. If determinism is unexpectedly shown to be true of this world, the two-pronged strategist believes that *compatibilism* would then be the right position to hold about agency and determinism – and the compatibilist about agency and determinism has no special problem endorsing compatibilism also about moral responsibility and determinism (although of course she *need* not do so, and might not want to do so for quite other reasons).

Moreover, it is a short step from this recognition to the conclusion that the two-pronged strategy permits the Agency Incompatibilist to be a Strawsonian optimist, rather than a pessimist. Recall that, for Strawson, the pessimist is a person who holds that 'if … [determinism] … is true, then the concepts of moral obligation and responsibility really have no application" (Strawson, 1962, p.72). But the two-pronged strategist does not fit this description. The first prong of her overall position is best put (as in (P2) above) by means of a subjunctive conditional: if determinism *were* true (as of course she believes it is not), there would be no agency. But as I shall go on to show in the next section, this is in fact perfectly compatible with thinking (second prong, and *indicative* conditional) that if determinism *is* true (as a matter of fact, as it were, and contrary to what she currently takes to be the case), agency remains an undeniable feature of reality. It is, admittedly, a further question, as just noted above, whether the two-pronged strategist would want to commit also to compatibilism about determinism and *moral responsibility*. It would be theoretically possible for a two-pronged strategist to endorse compatibilism about *agency* and determinism and yet to believe that moral responsibility required in addition the

[12] This is, of course, the corollary of the fact that the Agency Incompatibilist *does* accept (D2).

satisfaction of further conditions which could not be met in a deterministic scenario. The two-pronged strategist, then, *need* not be a Strawsonian optimist. But she also *could* be – and in this way, I contend, the way is open for the Agency Incompatibilist who takes the two-pronged route to avoid the charge of flying in the face of the Invulnerability Intuition.

(iv) Defending the two-pronged solution

It will be thought by many, no doubt, that there is something rather odd, or even straightforwardly illegitimate, about the two-pronged strategy. How can the Agency Incompatibilist both commit to incompatibilism about agency, and yet at the same time say that if determinism turned out to be true, there would still be agency? Isn't Agency Incompatibilism the claim that if determinism were true, there would be no agency? How, then, can the incompatibilist nevertheless maintain that if determinism *turned out* to be true, then she would maintain that there would be agency nevertheless? And even if these questions could be answered, in what sense is the resulting position any sort of *libertarian* means of hanging onto the Invulnerability Intuition, when it involves the admission that hanging onto it might involve having to become a compatibilist, should the Nightmare Scenario ever unfold?

The key to understanding this admittedly rather confusing-looking situation is to note that libertarianism in general (and so also Agency Incompatibilism in particular) consists of two separate claims: (i) the claim that there is free will/agency; and (ii) the claim that free will/agency is incompatible with determinism. But not all libertarians may take the same view of the relative credences it is rational to have in these two propositions. The Agency Incompatibilist who endorses the Invulnerability Intuition accords more credence to the first, believing it more securely known than incompatibilism itself. The reality of agency is the thing she believes there is most reason to insist upon[13] – and therefore in the event that she is forced to give up one of the two libertarian beliefs that is constitutive of libertarianism, because her epistemic situation is suddenly radically altered, she will choose to give up the idea that agency is incompatible with determinism. She will take the reasonable view that she must have been mistaken about that – even if she may not yet be able to see where exactly she went wrong.

This does not however mean we have to view her as a compatibilist *already.* She is not. She believes she knows there is agency and she believes that she also knows that if determinism were true, there could not be such a thing. She therefore thinks she *also* knows that determinism is *not* true, which implies of course that no scientists are going to come along to show that it is. As things stand, then, she is an

[13] Cf Samuel Johnson (1791): "Sir, we know our will is free, and there's an end on it" (p.80) To which he adds later: "you are surer that you can lift up your finger or not as you please, than you are of any conclusion from a deduction of reasoning" (p.273).

incompatibilist. That is the overall shape of the libertarian position I am envisaging. Is there something wrong with it?

Fischer has argued that there is. He believes that someone who has conceded that if scientists were ever to serve up incontrovertible proof of determinism, they would then become compatibilists is in fact required to "bring home" that rejection of incompatibilism – that is, to accept it *already*. In defence of this claim, he offers the following argument against Bailey and Seymour's contention that there is no such requirement – note that his specific target is van Inwagen and the argument is therefore based on the assumption that the relevant rational support for incompatibilism is supposed to be based on the Consequence Argument ('CA' in the quotation below):

> Van Inwagen claims that if he were convinced that causal determinism is true, he'd give up **Transfer.**[14] Given the generally accepted semantics for such conditionals and assuming van Inwagen's belief counts as knowledge, it follows that in the sphere of closest worlds to the actual world in which causal determinism is true, van Inwagen would give up **Transfer.** There is thus at least one possible world in which van Inwagen must suppose that Transfer is false. Because such principles have the status of "necessary", **Transfer** must not obtain in the actual world: if a proposition with this status is false in one possible world, it is false in all. You thus have to bring it (the rejection of **Transfer**) home and conclude that something (the invocation of **Transfer**) is actually wrong with CA. (2024: 200-01).

But this argument, I believe, is fallacious, as it stands. There is more than one issue here. For a start, the conditional Fischer is *explicitly* considering is this (as committed to by van Inwagen):

(VI) If I were convinced that determinism is true, I'd give up Transfer.

But this is a conditional about Van Inwagen's dispositions to believe things under changed circumstances, not about logical relations between determinism and Transfer. The semantics of conditionals merely implies that in the closest possible worlds in which Van Inwagen comes to be convinced that determinism is true, he gives up Transfer. But the fact that Van Inwagen gives up a proposition which is necessary, if true, in these worlds (and come to suppose it false instead) has no tendency to imply that he would give up that same proposition in any other worlds.

Perhaps, reading Fischer more charitably, the conditional he *really* means to be discussing is this: 'if causal determinism were true, Transfer would be false', which really *does* commit to relations between determinism and Transfer. But *in his present state of knowledge*, this conditional does not represent van Inwagen's position, and it

[14] 'Transfer' is one of the logical principles on which Van Inwagen's Consequence Argument for incompatibilism relies.

does not follow from (VI). At the present time, van Inwagen believes rather that if causal determinism were true, no one would be able to do anything other than what they do do – and hence free will would be impossible. He has no need to give up Transfer – more than that, he *should* not, since it is required for his argument for this present position. Van Inwagen's position is rather that if he came to know somehow that causal determinism *is* true (in the *actual* world, as it were), then retaining the most rational combination of beliefs in that situation would then involve ditching the Consequence Argument and its conclusion, by giving up Transfer. His commitment here is most charitably expressed, in my view, by an *indicative* conditional, not a subjunctive one: 'If determinism *is* true, then Transfer is false'. This conditional is like 'If Lee Harvey Oswald didn't kill Kennedy, somebody else did', not like 'If Lee Harvey Oswald hadn't killed Kennedy, somebody else would have'. And it is only the semantics of conditionals like the latter that are (more or less) uncontroversially given by the kind of possible worlds semantics that Fischer invokes. Note that I can perfectly well believe the former conditional without believing the latter. I may be really, really confident that Oswald killed Kennedy, but of course, it's not inconceivable that I'm wrong. So I put myself into the imaginary situation in which I know that I *am* wrong. And then I reason, that in this imaginary position in which I have somehow come to know that the killer wasn't Oswald, that in that case still, *someone* actually killed Kennedy after all, and if it wasn't Oswald, then presumably the shooter *must* have been someone else. But I'm not constrained by this belief to suppose that someone would have killed Kennedy even if Lee Harvey Oswald *hadn't*. I may (with good reason) judge this very doubtful, quite compatibly with hanging onto the truth of the *indicative* conditional. Fischer's argument, therefore, may be accused of relying on a point about the semantics of conditionals which may not apply at all to the best formulation of the conditional that van Inwagen should be taken to endorse.

It might be suggested that it is not clear, however, that this is the end of the matter. Helen Beebee has suggested to me (private correspondence) that the Oswald conditionals differ from the conditionals we are supposing van Inwagen might be endorsing in the following way. The evidence that might justify me in believing that 'if Oswald hadn't killed Kennedy, somebody else would have' is presumably *empirical* evidence. The truth of all the evidential propositions on which my belief is based is therefore perfectly consistent with my discovering that it is in fact rational to believe with a very high degree of certainty that the conclusion I have drawn from the evidence (the subjective conditional itself) is false and that (for example) the person I had quite reasonably suspected of wanting to kill Kennedy had no such intention – perhaps, for instance, I was basing my evidence on what I had supposed to be a diary entry, when in fact it was a piece of imaginative fiction. So though I might have to revise my conclusion, I needn't revise my premises. It's only become clear that my evidence was *incomplete* (as is always the case with empirical evidence) – so that my new knowledge that what I read was in fact a piece of fiction is able to change the answer to the question what it is now reasonable to conclude.

In the case of van Inwagen, however, what justifies his belief in the conditional which expresses his incompatibilism, that is, 'if determinism were true, there would be no free will', is a philosophical *argument*. So if he discovers that determinism is *actually* true, continuing to believe in free will requires him to abandon his philosophical argument for the subjunctive conditional. But he doesn't *thereby* come to have any *independent* evidence (independent, that is, of his now exceedingly strong evidence for supposing that determinism is true) that his reasoning process was flawed, or that it was missing a crucial element. His belief that that is so is *only* justified by the very high credence he now (but did not formerly) accord to the proposition that determinism is true.

I agree that this is a clear difference between the Oswald conditionals and the van Inwagen conditionals. Whether it is a difference that matters is less clear to me. Why exactly must the evidence which makes van Inwagen's shift of position rational be *independent* evidence? One possible answer might be that normally, if someone commits to a subjunctive conditional such as

(S1) If determinism were true, then there would be no free will;

then they normally commit thereby also to an associated conditional *about themselves*, such as:

(F1) If I were to *find out* that determinism was true, I would come to believe that there was no free will.

And then it might be argued that if van Inwagen is simultaneously claiming, along with (S1) that if he were to find out that determinism was true, he would *not* come to believe that there was no free will, this combination is inconsistent, and so he must already abandon one or the other.

But there are counterexamples to the claim that conditionals of the form of (S1) must always bring commitment to conditionals of the form of (F1) in their train. Consider, for example (S2) and (F2) below:

(S2) If the watery stuff in the rivers, seas, rain, etc. were XYZ, it wouldn't be water.

(F2) If I were to *find out* that the watery stuff in the rivers, seas, rain, etc. was XYZ, I would come to believe that it wasn't water.

Surely I can believe (S2) for the usual sorts of philosophical reasons which Kripke (1980), Putnam (1975), and many others have offered have for committing to such conditionals, without acceding also to (F2). It is much more plausible that if I were to find out that the watery stuff in my environment was XYZ, I would revise instead my idea of what water is and come to believe that it is XYZ instead. But that doesn't mean I can't commit *at present* to (S2).

It might be said that the possibility of committing to (S2) while demurring from (F2) is an unusual case, which depends on a certain indexicality which attaches to

the concept ‘water’ whereby its reference is fixed by whatever *in fact* is the nature of the watery stuff which surrounds me in the actual world. But it is not at all clear to me that one might not think of one’s fix on free will in a similarly indexical way - demonstratively, as it were. Whatever ‘free will’ is, we might imagine van Inwagen saying, it is whatever is going on *here* (offering paradigm cases of the exercise of free will) – and if these cases turn out not to involve any indeterminism, so much the worse for my incompatibilism. I do not think in fact that van Inwagen is in a good position to claim that he *does* think of free will in this way –what free will is, for van Inwagen, seems rather to be given by its association with the ‘could have done otherwise’ construction (or at any rate, that is what the form of his argument for its incompatibility with determinism suggests). But the Agency Incompatibilist, by contrast, is in a *very* good position to do so. On her view, agency is a distinctive biological phenomenon which is to be found represented throughout large swathes of the animal kingdom (just as water is a distinctive chemical substance) – and is therefore a phenomenon on which one might perfectly well take oneself to have a demonstrative kind of fix. It is true that she also takes it to be the case that this phenomenon, biological agency, is essentially indeterministic and that she has got to this position by means, mainly, of philosophical argument rather than scientific evidence. But this doesn’t seem to stand in the way of the Agency Incompatibilist’s insisting, still, that her way of fixing the reference of ‘agency’ is by means of demonstrative fix, rather than by *definitionally* associating it with indeterministic settling. Though she thinks that all agency *is* settling, she does not simply *define* it thus. And that is all that seems to be required to make it possible to insist that it *is* acceptable to hold both that agency is incompatible with determinism and yet at the same time that if determinism turned out to be true, agency (the demonstratively identified biological phenomenon) would still exist.

It seems to me, then, that Fischer has not established that it is illegitimate to ‘flip-flop’, particularly if one is an Agency Incompatibilist. In the final section of this paper, I want to go on to suggest that more than that, flipflopping is actually a better way to preserve the Invulnerability Intuition than Fischer’s own strategy.

(v) The Reasonableness of Optimistic Libertarianism

Two (rather partial) analogies may help to show why I think flipflopping is more appealing strategy for the defence of the Invulnerability Intuition than Fischer’s own resort to the metaphysics of semi-compatibilism. The first analogy comes from the very pure realm of *a priori* metaphysics and epistemology and concerns the choice between realism and idealism about the objects of perception. One might very reasonably think that we have an Invulnerability Intuition about the existence of tables, chairs, trees, flowers and the like. We might think that we know these kinds of objects exist – come what may – and might be confident that we will never – and should never - be moved from this belief. But – we may reason – if we are external

world realists about these objects, we can have at best only a shaky argument (perhaps e.g. an abductive one) for their existence, which would not really justify the degree of confidence we find we are inclined to have in their reality. In order to assert our right to be absolutely confident about them, it might be argued, then, that we ought instead to accept their ideality, in something like the way that Berkeley did.[15] Berkeley is insistent both that we are certain that "houses, rivers, mountains, trees, stones" exist and that by means of his principles, according to which these things are ideas, rather than mind-independent existents, "we are not deprived of any one thing in Nature" (p. 87). But most philosophers believe that the adoption of a Berkeleian view would be an over-reaction to philosophical scepticism[16] – and that, Berkeley's protestations notwithstanding, we end up (via this route) losing the essence of what we most wanted to defend. The thing we have the Invulnerability Intuition about, it turns out, is not merely the bare existence of these entities in some form or another – it is their existence *as things independent of our own minds* – a kind of existence we have to acknowledge we cannot absolutely prove. It is arguably far better, if we hope to defend a version of the Invulnerability Intuition truly worth the candle, to defend *as knowledge* the existence of mind-independent reality, while conceding, perhaps, that a person can never absolutely rule out analogues of the 'Nightmare Scenario' in which it is suddenly revealed that they have in fact been a brain in a vat their whole lifetime.

My suggestion is that the situation here is somewhat like the one we face in relation to our invulnerability intuition with respect to personhood and moral responsibility. Fischer chooses to restrict the scope of the intuition to encompass only the elements of personhood he regards as constitutive of moral responsibility – and not those constitutive of free will - in order to try to protect the intuition from the combined forces of powerful incompatibilist arguments together with the unfolding of a possible Nightmare Scenario. But as in the case of a potential retreat to Berkeleian idealism in order to avert the threat from scepticism, I would argue that this is to give up too much in the face of a threat which is merely theoretically conceivable. Better to try to protect a more full-blooded version of those intuitions – one which allows that we really are settlers of matters from moment to moment, as I believe it is undeniable we unreflectively believe we are, and not mere enactors of a pre-written script; better to defend the existence of real agency as the common-sense position, than to retreat to semi-compatibilism. Of course, since I have already conceded the conceivability of the Nightmare Scenario, I would have to acknowledge that just as it may turn out that I have been a brain in a vat my whole life, so it may turn out that determinism is true after all. But it is not rational to defend one's Invulnerability Intuitions in advance against these (in my view) equally implausible scenarios.

The difficulty I face with making this analogy at all persuasive to others is of course that, for what I believe are largely historical and sociological reasons, few

---

[15] See, for example, Berkeley (1710/1975), *Principles of Human Knowledge*.

[16] Not that Berkeley was himself motivated primarily by the desire to head off philosophical scepticism. His main concern was that the idea of material substance makes no sense.

people will accept that the truth of determinism is as unlikely as the brain-in-a-vat hypothesis to be true. For centuries, we have been encouraged to believe that the scientific truth about the universe is determinism – and so the need for a compatibilist strategy to save the Invulnerability Intuition doubtless seems much more compelling than an idealist framework to head off the threat of scepticism might. But I believe the consensus on this question is changing; and moreover, I believe that agency itself is a phenomenon that itself falsifies that old consensus which is staring us in the face, if we could only see it. There is not space here to defend any of these claims – but since Agency Incompatibilism is premised upon them, I hope it will at least be allowed that the *Agency Incompatibilist* may at least take the analogy to hold water.

The second analogy is more down-to-earth and comes from the highly applied science of engineering. If you live in an earthquake-prone zone, you nevertheless probably will not construct your buildings in such a way that they are likely to be able to withstand earthquakes of an intensity you take to be extremely unlikely ever to occur. You would be irrational to pay the price of the materials for a building as robust as this. Rather, the rational thing to do is to make your buildings as strong as needed to withstand any quake that has ever thus far occurred, perhaps with a certain margin added for additional safety. But if things change, of course, and evidence starts to reveal that earthquakes are gathering strength and becoming gradually more powerful as time goes by, you may later decide, in the light of this new evidence to move to more costly building materials. It is not rational *now*, though, to make this change, because the cost-benefit analysis does not currently support it. You will be paying an exorbitant additional cost to rule out a vanishingly unlikely scenario. And that is, by my lights, something like what Fischer is doing in adopting semi-compatibilism. The additional cost is failure to offer a robust defence of agency/free will (an essential component, for the libertarian, of the distinction between persons and non-persons) in order to rule out the possibility of having to capitulate in the face of a scientific proof of determinism, which is vanishingly unlikely ever to materialise.

For one who feels the force of the Invulnerability Intuition, then – and I certainly do –there are choices to make about how to defend it. Semi-compatibilism is one way to go – and is likely to appeal to those who already feel attracted to compatibilism, or to those who think that determinism has a reasonable chance of being true. But the libertarian may also consistently acknowledge its power – indeed she ought in my view to acknowledge the power of our Invulnerability Intuitions about freedom and agency also. It by no means follows from the mere conceivability of the Nightmare Scenario that the libertarian is out of options; and in certain respects, as I have argued, the route she takes in order to defend the Invulnerability Intuition has

considerable advantages over semi-compatibilism for anyone who thinks that free will is too large a prize to surrender for the sake of epistemological security.[17]

References

Bailey, A. and Seymour, A. (2021). In Defense of Flipflopping. *Synthese* 199: 13907-24.

Berkeley, George. (1710/1975). *Principles of Human Knowledge.* In M. Ayers (ed.) *Berkeley: Philosophical Works.* London: Dent.

Fischer, J.M. (1994). *The Metaphysics of Free Will*. Oxford: Blackwell.

Fischer, J.M. (2006). *My Way*. Oxford: Oxford University Press.

Fischer, J.M. (2016). Libertarianism and the Problem of Flipflopping. In Timpe, K. and Speak, D. (eds.), *Free Will and Theism*. Oxford: Oxford University Press. pp. 48-61.

Fischer, J.M. (2024). The Resilience of Moral Responsibility. In Cyr, T., Law, A. and Tognazzini, N.A. *Freedom, Responsibility and Value: Essays in Honor of John Martin Fischer.* New York: Routledge. pp. 189-210.

Fischer, J.M. and Ravizza, S.J. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Fischer, J.M., Kane, R., Pereboom, D. and Vargas, M. (2007). *Four Views on Free Will.* Malden, MA: Blackwell.

Frankfurt, H. (1982). The importance of what we care about. *Synthese* 53:2, pp.257-72, repr. in his *The Importance of What we Care About*. Cambridge: Cambridge University Press, 1998, pp. 80-94, to which page numbers refer.

Johnson, S. (1791), quoted by J. Boswell in *The Life of Samuel Johnson*, LL.D. (London: Penguin).

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Putnam, H. (1975). The Meaning of “Meaning”. *Minnesota Studies in the Philosophy of Science* 7:131-193, repr. in his *Mind Language and Reality*: *Philosophical Papers vol. 2.* Cambridge: Cambridge University Press. pp.215-71.

Steward, H.C. (2008). Moral Responsibility and the Irrelevance of Physics. *Journal of Ethics* 12, pp.129-45.

[17] I’d like to thank attendees at the Penelope Mackie Memorial Conference held at Nottingham University, 4-5 July 2024 for comments on some of the ideas in this paper. I’d also like to thank Helen Beebee, Kit Fine John Martin Fischer and Will Gamester for comments that have proven particularly helpful in refining my position.

Steward, H.C. (2009). Fairness, Agency and the Flicker of Freedom. *Noûs* 43:1, pp.64-93.

Steward, H.C. (2011). Moral Responsibility and the Concept of Agency. In Richard Swinburne (ed.), *Free Will and Modern Science* (Oxford: Oxford University Press), pp. 141-57; reprinted as chapter 36 of Dancy, J. and Sandis, C. (eds.), *Philosophy of Action: An Anthology*. Oxford: Blackwell, 2015.

Steward, H.C. (2012). *A Metaphysics for Freedom*. Oxford: Oxford University Press.

Steward, H.C. (2016). Libertarianism as a Naturalistic Position. In Timpe, K. and Speak, D. (eds.), *Free Will and Theism.* Oxford: OUP, pp.158-71.

Steward, H.C. (2021). What Is Determinism? Why We Should Ditch the Entailment Definition. In Marco Hausmann & Jörg Noller (eds.), *Free Will: Historical and Analytic Perspectives*. Springer Verlag. pp. 17-43.

Strawson, P.F. (1962). Freedom and Resentment. *Proceedings of the British Academy* 48, pp.1-25, reprinted in G. Watson (ed.) *Free Will.* Oxford: Oxford University Press, 2003, pp. 72-93, to which page numbers refer.

Van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Oxford University Press.