Steward, H. (2008) *Moral responsibility and the irrelevance of physics: Fischer's semi-compatibilism vs anti-fundamentalism.* The Journal of Ethics, 12 (2), pp. 129-145.
http://dx.doi.org/10.1007/s10892-008-9027-x

Moral Responsibility and the Irrelevance of Physics: Fischer's Semi-Compatibilism vs Anti-Fundamentalism

My Way is a collection of Fischer's recent work on moral responsibility which provides an excellent overview of the position that he has (with Mark Ravizza) steadily worked over the last twenty years or so to develop, clarify and defend – the doctrine he calls semi-compatibilism, which combines compatibilism about determinism and *moral responsibility*, with agnosticism about those more traditional varieties of compatibilism which assert that determinism is consistent with the existence of alternative possibilities. In some ways, indeed, *My Way* might be regarded as a better route into Fischer's distinctive outlook than *Responsibility and Control,*[1] the 1998 book, jointly authored with Mark Ravizza, in which the details of the semi-compatibilist theory are painstakingly worked out – for those details are complex, and it can be easy to lose sight of the shape of the semi-compatibilist woods while trying to make ones way amongst the very considerable number of trees which are required to stake out the territory. The essays which constitute *My Way* provide a somewhat gentler, though perhaps more circuitous route through the terrain. Early chapters provide an outline of the position which is less burdened by the considerable intricacies demanded by a book which aimed to be a complete statement of the 'official' version; while later ones, often written in response to specific worries or objections, seemed to me to offer more help with understanding how it is proposed to deal with the most obvious concerns one might have about the view, than does the rather denser presentation of *Responsibility and Control*.

[1] John Fischer and Mark Ravizza, *Responsibility and Control* (Cambridge: Cambridge University Press, 1998).

A brief sketch of the Fischer view on moral responsibility might go as follows. It is deeply implausible that our view of ourselves as morally responsible agents should depend for its sustainability upon abstruse considerations to do with the relation between physical laws and the events they subsume – whether, for instance, those laws (together with 'initial conditions') strictly entail the occurrence of those very events, or whether the laws are probabilistic only, leaving some leeway for alternative futures. But the question whether or not causal determinism is true just *is* the question what this relationship is between laws and initial conditions, on the one hand, and the events that occur in our universe, on the other. It would be a very good thing, therefore, if our account of moral responsibility were not to imply that it depends for its existence upon the falsity of causal determinism. But the arguments for thinking that causal determinism might be incompatible with the existence of *alternative possibilities* are rather strong (in particular, Fischer seems to think, some version of Van Inwagen's Consequence Argument might well show that determinism cannot coexist with alternative possibilities). It would be extremely helpful, therefore, if we were able to show that moral responsibility does not require the existence of alternative possibilities. Moreover, Fischer believes, there are no compelling reasons to think that moral responsibility is incompatible with causal determinism that do *not* depend on the assumption that moral responsibility requires alternative possibilities. If we were able to show that moral responsibility does not require alternative possibilities, therefore, we would be well on the way to carving out the required space for semi-compatibilism.

How might we show that moral responsibility does not require alternative possibilities? Fischer's strategy is to appeal to Frankfurt-style examples. Reflection on

these examples, he argues, reveals that what he calls *regulative control* – the sort of control that one only has if one could have done otherwise - is not a necessary condition of moral responsibility. All that is required is a less demanding variety of control, which he calls *guidance control.* But, Fischer contends, there is no reason to think that guidance control is incompatible with causal determinism. One has guidance control of one's action, roughly speaking, when one's action is produced by a mechanism which is "moderately reasons-responsive", provided only that the mechanism is "the agent's own". Detailed accounts are supplied of what it is for a mechanism to be moderately reasons-responsive, and also of what it is for a mechanism to "belong" to an agent in the relevant sense. Parallel accounts are also offered of moral responsibility for the consequences of one's actions, and for omissions. The whole constitutes an admirably thorough and careful development of a very distinctive variety of compatibilist thinking – compatibilist in the broad sense that it insists that our general view of *what we are* - persons, morally responsible agents, creatures with a certain sort of special significance - is compatible with causal determinism, even if we are sometimes inclined to claim powers for ourselves (e.g. the power to 'make a difference') to which, on reflection, we might have to renounce our claim, in the event that causal determinism turns out to be true.

For someone who (like me) is an incompatibilist about moral responsibility and causal determinism, there are several avenues of response available to this semi-compatibilist line. One might try to insist that the Frankfurt cases do not show what Fischer takes them to show, and that alternative possibilities *are* necessary for moral responsibility after all. Or one might try to argue (*contra* Fischer) that even if alternative possibilities are *not* required for moral responsibility, there are good reasons of some

other sort for supposing that moral responsibility is not compatible with causal determinism (for example, that we could not be the ultimate originators of anything unless causal determinism were false, and that moral responsibility requires this power of ultimate origination). Or one might try to combine responses of both these kinds – for one might quite reasonably think (as has been argued, for example, by Robert Kane)[2] that there is more than one reason why moral responsibility is incompatible with causal determinism. But such replies as these are bound ultimately to fail to convince, so it seems to me, unless they are backed up with some indication of what the incompatibilist is to say about the powerful claim which is clearly Fischer's starting point, and which provides his motivation for supposing that some theory with roughly the shape of semi-compatibilism is likely to be right – that is, the claim that our status as morally responsible agents ought not to depend on what Fischer calls "the arcane ruminations – and deliverances – of the theoretical physicists and cosmologists" (p.5) For that point, I concede, seems absolutely correct. It just seems preposterous to suppose that anything a physicist might report as the ultimate scientific verdict about the nature of the basic physical laws which govern the universe, might have the potential to settle the question whether or not we truly are morally responsible agents. As Fischer says, "Our fundamental nature as free, morally responsible agents should not depend on whether the pertinent regularities identified by the physicists have associated with them (objective) probabilities of 100% (causal determinism) or, say, 98% (causal indeterminism)" (p. 5). And unless this point is properly addressed by the incompatibilist, I suggest, nothing she has to say about Frankfurt-style scenarios or ultimate responsibility is likely to be completely satisfying. My first objective in this article, therefore, will be to explain why I

[2] See *The Significance of Free Will* (Oxford: Oxford University Press, 1996).

believe it is possible for an incompatibilist to side with Fischer on the irrelevance of basic physics to the question whether or not we are morally responsible agents – and to sketch the shape of the picture of action, causality and levels of explanation which would be required to secure the availability of the wanted position for the libertarian. I shall then return, in the second half of the article, to consider what might be said specifically about Fischer's claim that Frankfurt-style examples can be used to show that moral responsibility does not require alternative possibilities. My suggestion will be that Fischer's failure to see that there is room for an incompatibilist to *agree* with him about the irrelevance of physics to moral responsibility, is related to his inability to discern a possible ground for supposing that alternative possibilities are necessary for moral responsibility which the Frankfurt cases leave untouched.

(i) *Moral Responsibility and Basic Physics*

Let me begin by quoting a little more fully from the paragraph in which Fischer first explains the importance for his thinking of the idea that our judgements about moral responsibility should be (what he calls) *resilient*:

> …. our basic status as distinctively free and morally responsible agents should not depend on the arcane ruminations – and deliverances – of the theoretical physicists and cosmologists. That is, I do not think our status as morally responsible persons should depend on whether or not causal determinism is true. (p.5)

I accept the claim made by the first of these sentences. But what I should like to ask is whether that must entail accepting the second – whether the casual 'that is', with which Fischer joins these sentences together, is really justified. Must we accept that whether or not causal determinism is true is a question which must be put in the hands

of "theoretical physicists and cosmologists"?[3] If it is, then evidently enough, an incompatibilist must accept that it is ultimately up to the theoretical physicists and cosmologists to tell us whether or not it is possible that we are morally responsible agents. But perhaps there is room for manoeuvre in the recognition that despite that nonchalant 'that is', Fischer makes a *step* in his reasoning here, one which the incompatibilist might think to question.

If we are to decide who might have the right to tell us whether the doctrine of causal determinism is true or not, we are going to have to know what that doctrine asserts. Fischer claims that

> the doctrine of causal determinism entails … that for any given time, a complete statement of the (temporally genuine or non-relational) facts about that time, together with a complete statement of the laws of nature, entails every truth as to what happens after that time. (p.5)

Let us take this, then, for the time being, as our guide to the question what the thesis of causal determinism says. Is it up to physicists to tell us whether or not it is true? It is not *immediately* obvious why it should be. On the face of it, there seems no reason why, amongst the "temporally genuine facts" we might not find included (for example) certain biological, psychological, sociological, and economic facts, about which physicists can claim to have no particular expertise – not to mention a whole pile of utterly mundane particularities which belong neither to any scientific nor indeed to any other domain of enquiry – such as that there is currently a globe on my

[3] It is not immediately obvious to me who Fischer has in mind when he refers here to 'cosmologists'. In a sense, of course, anyone concerned with extremely general questions about the nature of the universe as a whole could be regarded as a 'cosmologist' – and so it is almost a definitional truth that anyone concerned with a doctrine of such great generality as the doctrine of causal determinism would count as a cosmologist. But I suspect that Fischer is really thinking of cosmology here as a particular branch of *physics* – the branch which is concerned with such matters as the origin and size of the universe, the nature of the matter it contains and of the physical laws which it instantiates. I shall assume in what follows, therefore, that his point is simply that it should not be within the remit of *physics* to settle the question whether or not we are morally responsible agents.

desk, that my printer is out of ink, that there are no elephants in this room and that my favourite colour is green. Neither have we been told why the "laws of nature" are to be regarded as entirely the preserve of the physicist; it is not immediately obvious (again, without further argument) why some of them might not belong to geology or chemistry or biology; or even to psychology or economics or sociology. *As stated*, then, there is no immediate connection between the thesis of causal determinism and physics – that connection is going to have to be *forged*.

It is not difficult, though, to see how it is that the thought arises that physics might be the science uniquely well-placed to judge of the question whether causal determinism is true**.** For the thesis of causal determinism, as Fischer describes it, is dependent for its very coherence on our being able to make sense of the suggestion that there might be such a thing as the "complete statement" of the "temporally genuine" facts about any given time. Not only do we need a workable suggestion as to how we are to distinguish a "temporally genuine" fact from a relational one, in such a way that we will be able to understand clearly which facts concerning the present fall into the set from which the thesis of causal determinism alleges it is in principle possible to infer those concerning the future. We also need to be confident that the idea of there perhaps being a "complete statement" of the facts about any given time is not one of whose undermining we are going to be able to make extremely short philosophical work. And this in turn would seem to require that there be a satisfactory way of *constraining* those facts, so that, for example, the fact that there are fewer than 9065 discrete objects with a mass greater than 1g on my desk at present does not count as one of the facts in question (for if *that* counted, then presumably, the fact that there

are fewer than 9066 such objects, and fewer than 9067 … and so on, would also have to count, and we would be faced with an infinite number of facts, which seems difficult to square with the idea that there might be such a thing as a "complete statement" of them). It is reflection on how we might attempt to deal with such difficulties as this that ultimately generates the thought that it is physics which must, in the end, answer the question whether or not causal determinism is true. For the idea to which we are going to have recourse, presumably, will be the thought that there is a finite set of *basic* facts (perhaps facts, for instance, about the nature and distribution of fundamental physical particles) from which such further facts as those about the number of objects on my desk weighing more than 1g, might be simply inferred, so that we need only mention these basic facts in our "complete statement", leaving the others to follow inexorably. It is *this* idea which must motivate the thought that physicists, in particular, might have some particular expertise to bring to bear on the question whether causal determinism is true. If all the facts, however complex, depend ultimately on the properties of basic physical particles and the laws which they follow, then it would seem that we have to defer to physicists when considering the question how the world might conceivably unfold. It is they who will tell us whether or not causal determinism is true, for it is they who will tell us whether the laws governing these basic physical particles are deterministic, or whether they are probabilistic only, leaving scope for multiple unfoldings.

But the question is whether everything, at every level, really *does* depend in the way envisaged by this picture on the properties of basic physical particles and the laws which they follow. Philosophers are often encouraged to think that believing in

the "supervenience" of higher level facts on basic physical ones is a minimum requirement for naturalistic sanity in metaphysics. But it is not obvious that supervenience, as generally characterised, entails that the evolution of reality over time depends (in so far as it depends on anything) only on physical laws. What supervenience on the physical is usually said to dictate is that no two worlds which were identical in all basic physical respects at a given time *t* could differ in any supervenient respect at that same time *t*. And one might conceivably concede *that*, I think, while still refusing to accept that the question what makes reality evolve as it does *over time* is a matter only for physicists. Even if one adds to the characterisation of supervenience the suggestion that there may not be a *change* in any supervenient property without a change in the 'base' properties on which it supervenes, we have not yet arrived at the idea that any given supervenient change must *depend for its explanation* upon the subvening one which (if we concede the supervenience of a wide range of higher-level properties and changes on physical properties and changes) must always accompany it. The direction of explanation might (in some instances) rather be the other way about. So far as I can see, it seems to be consistent with the supervenience of the "temporally genuine" higher level facts on the lower level physical ones, that physical laws are not (even in principle) sufficient to determine the state of the world at t2 given its state at t1 – and not just for the standardly-touted reason that perhaps some of those physical laws will ultimately be confirmed by physicists to be probabilistic. One might just think (as does Nancy Cartwright, for example)[4] that such physical laws as there are, are narrow in their scope, strictly applicable only to the relatively small number of situations that correspond closely to

[4] See her *The Dappled World* (Cambridge: Cambridge University Press, 1999).

the physical models which supply those laws with their concrete interpretations, and applying only *ceteris paribus*, even where they do apply. Indeed, in some ways, the idea that the evolution of reality over time might depend solely on 'initial conditions', together with purely *physical* laws, is a quite extraordinary one. And I do not mean, in saying this, to imply that it is extraordinary to believe in physical-causal determinism as opposed to physical-causal indeterminism. I mean that it is extraordinary to believe *either* that laws at the lower level entirely *determine* (as opposed to *constrain*) the development of reality *or* that these laws "fix the probabilities" concerning how reality will evolve, leaving some scope for chance. After all, the evolution of reality is profoundly influenced (we tend to think) on a large scale by such things as wars, stock market crashes, global warming, revolutions, industrialisation, etc., as well as (on a small scale) by the myriad small decisions each of us makes on a daily basis. To suppose that the occurrence of any of these sorts of things is no more than the high-level manifestation of the inevitable workings-out of the consequences of the initial conditions at the start of the universe (determinism) – or else of those initial conditions and merely probabilistic laws, together with a smattering of chance (indeterminism) - is perhaps one of the most astounding things that has ever managed to obtain the status of philosophical orthodoxy (though it must be conceded that there is strong competition for this title). To believe this would seem to be to consign all sorts of factors which it is natural to regard as causally crucial to the realms of the utterly epiphenomenal. Nothing really matters, it would appear, in anything other than an extremely attenuated sense of 'matters', to the unfolding of the world, except the way physical reality was in the beginning, the physical laws and (perhaps) chance. How are

we to make room, given this picture, for our basic conviction that *we* matter to that unfolding – both individually, through our actions, and as a species, through the phenomena to which our activities have given rise – societies, governments, armies, businesses, religions, technologies, art, literature, science?

There are, of course, many attempts to show that our mattering in all these ways is perfectly *compatible* with everything's nevertheless being fixed by what goes on at the level of the smallest constituents of reality. And I cannot here attempt to show that these manoeuvres are bound to be unsuccessful (though I believe that to be the case). My point here is fortunately less ambitious than that; I aim only to show that there is certainly dialectical space for a certain kind of incompatibilist to *concede* Fischer's point that it is unacceptable to suppose that the question whether any of us is really morally responsible for anything is a matter which physicists will ultimately have to settle. For the incompatibilist claims only that *causal determinism* is incompatible with moral responsibility. She need not accept that whether or not causal determinism is true is a question merely about the nature of the basic physical laws. *That* only follows if one accepts a certain picture of the relationship between the various 'levels' of ontology, explanation and causality - a 'bottom up' picture whereby the evolution of reality over time is conceded to be determined entirely by physical laws (perhaps together with some admixture of chance, if these laws happen to be probabilistic only). My suggestion is that *this* is where the incompatibilist should demur. She should insist that there is more than one way in which we can conceive of causal determinism's being false. It might be false (as Fischer imagines) because, although it is true that (i) the evolution of reality over time depends entirely

on the evolution of *physical* reality over time, as dictated by purely physical laws, nevertheless (ii) those physical laws are indeterministic. But it might also be false because (i) it is *not* true that the evolution of reality over time depends entirely on the evolution of physical reality over time, as dictated by purely physical laws and (ii) there is, moreover, no *other* set of laws of nature which, together with the facts at a given time t, might be thought to entail the facts at t + 1. And the availability of this second means of arguing for the falsity of causal determinism shows that the question whether or not it is true is not necessarily a physicist's question. It only turns into a physicist's question once we decide to return a certain answer to the prior *metaphysician's* question how we are to conceive the relation between what we are used to thinking of, metaphorically, as the *levels* of reality. The incompatibilist, I suggest, ought not to return the answer which delivers the question to the physicists to settle.

It is not, after all, as though the belief that physical laws neither entirely determine the future, nor fix the chances that it will evolve globally in any given way, would be an *ad hoc* resource for an incompatibilist to exploit. For what is at the heart of the incompatibilist's thinking, of course, is the idea that on occasion, a *human agent* can determine how the future will be in certain respects. That thought sits ill with the idea that physical laws, together with initial conditions at the start of the universe, determine the entire future of that universe (which is, of course, what gives the incompatibilist the basic motivation for her incompatibilism). But, as compatibilists never tire of pointing out, that thought *also* sits ill with the idea that physical laws *fix the chances* that a certain future will evolve. What the

incompatibilist appears to need is a view which releases us from puppet-like enslavement either to deterministic *or* to indeterministic laws of physics. And that, I suggest, is not going to be had on the metaphysical cheap. It requires no less than an re-examination of various doctrines concerning the basicness of the physical and the universal grip of physical law which we have grown used to thinking of as non-negotiable.

It might perhaps be responded by Fischer that it is not much better if our basic status as distinctively free and morally responsible agents turns out to depend on the arcane ruminations and deliverances of *metaphysicians*, than that that status should depend on the verdicts of the consortium of scientists at Caltech, Stanford and MIT whom he at one point imagines announcing the news that (contrary to previous received opinion) the equations which govern the universe have turned out to be deterministic after all. But it *is* better – much better. A metaphysician's ruminations concerning such matters as the reach of law, the nature of causality, the relationship between levels of explanation, etc., must *include* considerations deriving from such things as the all-but-indisputable fact of agency; the apparent efficacy of factors of a social, economic and psychological kind; the role which appears to be played by *reasons* in the generation of human (and other animal) activity - indeed, these very things constitute the facts to which she must attempt to make her theories answerable. What worries Fischer about views on which moral responsibility depends upon the falsity of causal determinism, is the prospect that an expert verdict might be handed down from scientists concerning matters which it is indisputably entirely within the scope of their discipline to determine – and that we would then just have to live with

the consequences for moral responsibility. I agree that this picture of the relationship between physics and moral responsibility is utterly untenable. But there is more than one way to reject it. One way is to be a semi-compatibilist. Another is to reject the *fundamentalism*[5] which would hand such power to the discipline of physics in the first place.

It might be insisted that we still do not *know* whether causal determinism is true, even if, as I have suggested, it is not necessarily down to physicists to tell us whether or not it is. And it might be suggested that our status as morally responsible agents ought not to depend on the falsity of a thesis we cannot, at present, *know* to be false. But in the new context which is provided by the anti-fundamentalist position, this claim seems unreasonable. The anti-fundamentalist incompatibilist believes that the apparent existence of free agency (and perhaps other considerations, too) provides strong reasons for rejecting the thesis of causal determinism, which she insists must be regarded as a *metaphysical* thesis, not a scientific claim. Perhaps she must concede (as must anyone with a modicum of modesty) that she cannot know for sure that she is right about this – perhaps it might turn out in the end that she is mistaken, and that causal determinism is true after all. But it can hardly be insisted that it is a condition of offering a case for believing some proposition *p* that one know for sure in advance that one's case is watertight. (Of course, it must not be the case that one knows that one *has* made such mistakes, but that is quite a different matter). What the anti-fundamentalist incompatibilist maintains is that her argument for incompatibilism, together with what appear to be the facts about free agency, justify her insistence that causal determinism is false. And since she does not suppose that causal determinism

[5] I borrow this term from Nancy Cartwright, *op cit.*

is merely a claim about the nature of physical law, this is not an *a prioristic* punt on the deliverances of future scientific theorising, such as Fischer supposes that the libertarian must inevitably take. Free agency is not thought of by the non-fundamentalist incompatibilist as dependent on some (as it were) *subterranean* truth about the ultimate workings of the world. Free agency, on her view, just makes causal determinism false, straightforwardly and directly. It is a *counterexample* to the thesis. Where it occurs, she insists, something happens in the world (something which, according to her preference, she might call a decision, a choice, a volition or an action) which was *not* determined by the conditions at any preceding time together with the laws of nature. And though she concedes that it might turn out that she is wrong about this (on general grounds of fallibilism), she does not think she needs to wait on *physics* for the confirmation of her verdict. For physics, on her view, contains no laws of the all-embracing, world-dominating sort that might rule out the possibility of such a thing as a causally undetermined decision, choice or action. Physics is just that – physics – and in the anti-fundamentalist's opinion, it has nothing much to say, and never will have, about the provenance of actions.

Nevertheless, even if logical space is available for an incompatibilism which does not concede to physics the ultimate verdict on the question whether or not we are morally responsible agents, the view I have sketched is likely to be regarded a radical view. It requires that we reject a 'bottom up' picture of reality with which many philosophers have grown comfortable and to which it might be difficult to work out a coherent alternative. It might demand that we be able to make sense of notions that philosophers have struggled to clarify and often rejected out of hand – such as

'downward causation', for instance. Surely semi-compatibilism is the better option! I want now to try to explain why I do not think this is the case – and why, therefore, I believe that anti-fundamentalism, and not semi-compatibilism, is the way to go for someone who believes, as Fischer and I both do, that the question whether we are morally responsible agents is not one which is ever going to be settled by physics.

(ii) *The Importance of Alternative Possibilities*

Semi-compatibilism rests, ultimately, on the idea that we can give an account of the powers distinctive of agency without recourse to the idea that an agent must have real access, at the very moment of her action, to alternate possibilities. Fischer relies for his explication of how this is possible on the contrast between what he calls *regulative* control, on the one hand, and *guidance control* on the other. Moral responsibility, he alleges, requires only guidance control; and he illustrates the distinction between the two kinds of control with an example:

> … moral responsibility does not require regulative control. To see this, suppose you are at the controls of an airplane, a glider, and you are guiding the plane to the west. Everything is going just as you want, and the plane is making good headway. You consider whether to steer the plane to the east, but you decide to keep guiding it to the west, in part because the scenery is nicer in the west. Unknown to you, the wind currents in the area are such that the plane would continue to go to the west, in just the way it actually goes, even if you had tried to steer it in some other direction. … In this example, you steer the plane to the west in the "normal" way. It is not just that you cause it to go to the west (which you would equally have done had you steered the plane in the same way as the result of a sneeze or an epileptic seizure). Rather, you guide the plane in a distinctive way – you exhibit a signature sort of control which I shall call "guidance control". Here, you exhibit guidance control of the plane's movements, but you do not possess regulative control *over* the plane's movements (p.8).

What is the role of this example supposed to be? It cannot be intended as a straightforward *counterexample* to the claim that moral responsibility requires

regulative control – as might perhaps be confusingly suggested by the way in which Fischer introduces it (“ … moral responsibility does not require regulative control. To see this …”). For it does not follow from the fact that I do not possess regulative control over the plane’s movements under the circumstances envisaged that I do not possess regulative control, in these same circumstances, over a great many *other* things – notably, perhaps, in this case, over the movements of my own body. I can push levers this way and that; press buttons (or not); even desert the cockpit altogether and leave the plane to its own devices. And it might be that my moral responsibility for the plane’s movements ultimately depends in this case on my possession of regulative control over these *other* matters. Fischer himself, indeed, seems to recognise that the case does not succeed in proving that moral responsibility does not require regulative control – for after presenting it, together with another example of Locke’s, he asks: “Do such examples show that one can be morally responsible for some behavior, even though one lacks freedom to choose or do otherwise, that is, lacks regulative control?” And he responds thus: “The problem is that, apart from any special assumptions, such as causal determinism, it is plausible to suppose that you could have chosen to steer the plane in a different direction, tried to do so, pushed the steering apparatus in a different way, and so forth” (p.9). It appears, then, that Fischer himself does not regard the plane case as a counterexample to the claim that moral responsibility requires regulative control. The idea must be, then, that it is supposed to be simply an *illustration* of how it might be possible to have one kind of control (guidance control) over something (in this case, the plane’s

movements), while lacking control of another kind (regulative control) over that same thing.

It seems reasonable to suppose that the example really does illustrate that one can exercise a certain sort of control (the sort that Fischer calls guidance control) with respect to the movements of a certain thing without it being within one's power, due to certain counterfactual truths, to alter those movements in any way. But if moral responsibility is to be compatible with causal determinism, and we are to work, as Fischer wishes us to, on the assumption that causal determinism rules out alternative possibilities, it must be shown that an agent's being morally responsible for something is compatible with her lacking regulative control not just over that same something, but with her lacking regulative control over *anything whatever*. If the universe is causally deterministic, no agent ever has any regulative control over anything, for regulative control is defined as a kind of control which requires genuine access to alternative possibilities. As Fischer notes, "causal determinism would extinguish not just a prairie fire of freedom, but also the tiniest flicker" (p.41). And the question arises, therefore, whether moral responsibility can exist in the complete absence of the phenomenon Fischer calls regulative control. This is what must be shown by the semi-compatibilist who wishes to carry conviction.

How might this be shown? The trouble is, as Fischer is well aware, that it is well nigh impossible to construct a case based on the sort of counterfactual scenario that is imagined in the airplane case ("If I had tried to guide the plane to the east, it would still have gone to the west") which has the structure Fischer really needs for his argument – a case in which an agent acts, and in which it seems intuitively right to

regard her as morally responsible for what she does, but in which *all* alternative possibilities whatever are blocked by the truth of certain counterfactual propositions. Admittedly, *most* alternative possibilities are ruled out by the sorts of cases which have come to be known as 'Frankfurt-style' examples – and it is indeed to these cases that Fischer looks for the basis of his argument. In a Frankfurt-style example, we are asked to imagine an agent (call him Jones) whose thought processes are constantly monitored by a second agent (call him Black), who wants Jones to act in a particular way – say, to vote for Clinton. Should Jones show any signs of being tempted not to vote for Clinton (let us suppose, for example, that he would blush just prior to the onset of any decision to vote for another candidate, and that Black knows this), Black would intervene neurologically, to ensure that Jones votes for Clinton in any case. But in the event, there is no need for an intervention on Black's part. Jones just freely votes for Clinton in the normal way. We are invited to agree that in such a case, Jones is morally responsible for his action. And we are invited to agree also that he does not, under such circumstances, have regulative control over his action – since, given Black's presence, the structure of his motivations and his powers, Jones could not have done other than he did. But Fischer concedes that even under such unfavourable circumstances as are here envisaged, *certain* alternative possibilities appear to remain open to Jones. Jones could, for example, have shown a sign that he was about to be tempted to decide not to act in the way desired by Black (thus triggering Black's intervention) or he could, perhaps, have refrained from acting in the wanted way *on his own*. Frankfurt-style cases of this variety cannot be used *straightforwardly,* then, to show that it is possible for an agent to be morally responsible for what she does

even though no alternative possibilities are available to her at all. Fischer's strategy is instead to argue that at any rate, these Frankfurt-style cases rule out all alternative possibilities which might conceivably be *relevant* to the question whether or not the agent is morally responsible. He insists that it is completely implausible to suppose that any of the alternative possibilities which remain open to the agent in a Frankfurt-style case could be the sorts of alternative possibilities which might *ground* the judgement that the agent was morally responsible for the action he in fact undertakes. The alternative possibilities in question, he insists, are mere 'flickers' of freedom, which are insufficiently robust to make the difference between a case in which an agent is, and one in which she is not, judged morally responsible for what she does. And given that that is so, he concludes, it is reasonable to maintain that the Frankfurt cases do indeed show that moral responsibility does not depend upon alternative possibilities.

But not all the alternative possibilities that Fischer is inclined to dismiss as mere 'flickers' of freedom seem equally unimportant. Let us consider some of them. One is the alternative that is represented by the possibility that Jones might have blushed (thereby triggering Black's intervention). I agree with Fischer that it is most implausible that the incompatibilist about determinism and moral responsibility is going to be able to make anything much of the existence of this alternative possibility. The reason is this: that any plausible, non-question-begging principle relating moral responsibility to the existence of alternative possibilities is surely going to have to relate to the need for the agent to have certain *powers* if she is to be held morally responsible for what she does. And even if a certain openness is present in the

Frankfurt case envisaged, in virtue of which we are inclined to say that Jones might have blushed, thereby triggering Black's response, this does not seem to be an openness that is attributable, in any way, to a power possessed by Jones. Blushing, as it is supposed to figure in the example, is meant to be conceived of as an involuntary business – it isn't up to Jones whether he blushes or not – and therefore it seems implausible to suppose that the possibility that he might have done could help to show that he is morally responsible for voting for Clinton.

But not all of the alternative possibilities which Fischer writes off as mere 'flickers' are like this. One possibility which seems to be open to Jones, for example, and which genuinely does appear to constitute a *power* of his, is this: he is able not to perform the particular action of voting for Clinton that he does in fact perform. It is very plausible that when Black intervenes, what he produces cannot be an event identical with the action that Jones actually undertakes when he freely votes for Clinton. Indeed, it might be argued that Jones has it in his power not to perform an *action* of voting for Clinton at all – for one might reasonably ask what reason there is to suppose that a chain of events which is initiated by Black might count, nevertheless, as an *action* on the part of Jones. Fischer's response to some rather similar suggestions that have been made by others in the literature leads me to believe that he would respond to this suggestion by saying that such possibilities as these remain too slender, too exiguous, to be relevant to the question of moral responsibility. But it seems to me that Fischer is wrong about this, for a reason I shall now attempt to explain.[6]

[6] I give a fuller account of this argument in 'Fairness, Agency and the Flicker of Freedom', forthcoming ??.

In assessing the question just how 'robust' an alternative possibility has to be before it is relevant to the question whether or not an agent is morally responsible for what she does, I think it is essential to distinguish between two quite different sorts of reason one might have for supposing that moral responsibility might require alternative possibilities. Many philosophers who have supposed that alternative possibilities are required for moral responsibility assume that the connection depends on a principle, essentially, of *fairness* - the basic intuition being that it is simply not *fair* to hold an agent morally responsible for φ-ing, say, when she simply could not have done other than φ. And if *this* is the basis on which one hopes to justify one's belief in the claim that alternate possibilities are necessary for moral responsibility, it might indeed seem as though the sorts of possibilities which I have alleged might be available to Jones are much too flimsy to be relevant to the question whether or not he may be held morally responsible for voting for Clinton. For the power not to perform the *particular* action one undertakes seems often to be available to agents of just the sort that one might hope the appropriate fairness principle would *excuse*. A heroin addict, for instance, who injects heroin into her right arm at *t* might have been able to wait a couple of seconds and inject the heroin at *t + 2* instead; or might have been able to inject methadone instead of heroin; or might have been able to inject the heroin into her left arm rather than her right; and in virtue of these powers, might be said to have had the power, at *t*, to refrain from performing the *particular* action she did in fact perform. But how could such apparently unimportant little powers as these give her the sort of control which would make it fair for us to judge her morally responsible for her action? Such a suggestion would surely be ludicrous. It is perhaps

not clear to what extent a heroin addict ought to be judged morally responsible for a given act of heroin injection – but it is surely clear enough, it might be said, that the answer does not depend on the availability of such exiguous possibilities as those mentioned above.

But one need not suppose that it is a principle of *fairness* which most crucially relates moral responsibility to alternative possibilities. One might rather think that the deepest and most important reason why alternative possibilities are required for moral responsibility is that they are required for *agency* (agency being, in its turn, a necessary condition of moral responsibility). For one might believe (particularly, of course, if one is a libertarian) that actions are events of a very special sort. Properly understood, the libertarian might insist, bodily actions are not mere bodily movements – rather, they are the causings of bodily movements *by agents*.[7] And in order for a movement to count as having been caused by an agent (as opposed to a mere process internal to that agent's body), she may say, that movement will have to have been the product of an exercise of what Reid called a *two-way* power on the part of that agent – to act, or to refrain from acting – so that in the absence of the relevant power of refrainment a genuine action simply could not occur. And for someone who holds this view, the power of refrainment that is crucial to the question whether an agent who φ-s on a given occasion in fact acted on that occasion is not the power not to have φ-ed (as indeed Frankfurt-style cases show rather clearly – for we retain the strong intuition in those cases that the agent acted, *despite* lacking the power not to have φ-ed). The

[7] For simplicity of exposition, I ignore purely 'mental' actions – such as conjuring up a mental image of the Eiffel Tower, or adding 23 and 49 in one's head - though it seems to me not impossible that one might think of arguing that these actions, too, are essentially constituted by an agent's causing certain bodily events (presumably ones that occur, in this case, in the brain).

power of refrainment that is relevant seems rather to be a power that is, as it were, internal to the particular action which was in fact performed – the power not to have performed *it* – a power that could have been, but in the event was not, exercised by the agent at the time of the action. The existence of this power, on this view, is what makes it correct to say that the action *was* an action – and in the libertarian's opinion, it will always be crucial to what Fischer calls 'the signature sort of control' that we find to be present in the activities of the pilot guiding the plane to the west. Where an agent lacks the power to refrain from the particular φ-ing she in fact undertakes, on this view, her φ-ing on that occasion cannot be thought of as an exercise of agency at all – it is a mere event, over the occurrence of which it is inconceivable that the agent should have any control. And against the background of this line of thinking, Fischer's claim that the power to refrain from undertaking a particular course of action is too exiguous to 'ground' moral responsibility loses its power. For on this view, the power to refrain is relevant in a way different from that imagined by Fischer – it is not supposed to 'ground' moral responsibility directly, by supplying a condition that any morally responsible agent must meet if it is to be *fair* that she is held morally responsible for what she has done. Rather, it is conceived of simply as a necessary condition of any genuine exercise of agency. It is a necessary condition, moreover, that Frankfurt-style agents appear to meet. But the incompatibilist will want to insist that it is not a necessary condition that could be met if causal determinism were true. And therefore she will disagree with Fischer's verdict that Frankfurt-style cases can be used to show that alternative possibilities are not necessary for moral responsibility. For Frankfurt-style agents *retain* what she regards as the most crucial

alternative-possibility-involving power there is – which is simply the (two-way) power to *act*.

It might be retorted that it is simply not plausible that an action is, by its very nature, a kind of event which is inconsistent with the truth of causal determinism. I want next to consider three possible arguments for this claim. I shall argue that none is satisfactory – and that all are, in a way, connected with the failure to take seriously the possibility that is made available by the anti-fundamentalist line I considered in section (i).

(iii) *The Nature of Actions*

Fischer assumes, as do many participants in the contemporary debate about free will (including many incompatibilist participants) that there is nothing about action *per se* which is inconsistent with causal determinism. But this has always seemed to me a strange assumption. When I imagine the universe as it is alleged to be by the causal determinist, a place in which each event follows inexorably from immediately preceding circumstances according to all-embracing laws of nature, it seems to me that that universe permits no space for the phenomenon of action at all. It is a world of mere events in which there seems to be no place of the sort where an agent might interpose herself and make a difference to what then occurs. The compatibilist, doubtless, will want to argue that the idea that an agent needs such a 'place' to interpose herself into the series of events is an unnecessary and possibly even incoherent one – that an agent's interpositions must simply form a special subset of the networks of deterministically related events which she envisages as constituting

causal reality. But that the suggestion is largely absent even from contemporary presentations of *incompatibilist* arguments seems to me curious. Even agent causationists, who are likely to agree that one cannot think of what they call the *free* actions of an agent merely as events which occur either inexorably, or with a certain determinate degree of probability, in consequence of previous ones, often concede that there are, nevertheless, many actions of lesser importance which are indeed merely the deterministic consequences of prior conditions. To concede this, though, is to cut oneself off, so it seems to me, from a powerful idea – the idea, that is, that actions might be metaphysically distinctive simply in virtue of being actions. It is to cut oneself off from the prospect of a *unified* account of actions, which might relate the role of the agent, the concept of a two-way power, and the correspondingly distinctive profile of the events which are actions, in an illuminating way. Why, then, have philosophers been so reluctant to entertain the possibility that actions *per se* might be inconsistent with causal determinism?

There are, I think, at least three important factors which have conspired to prevent many contemporary philosophers (Fischer included) from seeing that a challenge to compatibilism might be mounted on the basis of the simple claim that causal determinism is inconsistent with agency. Of these, the most important, I think, is the assumption which I have already challenged in the first part of this paper, that causal determinism is a doctrine which might, for all we know, be true. Given that assumption, the suggestion that agency and causal determinism are incompatible is likely to seem preposterous. Surely, it might be said, we cannot accept a view according to which, for all we know, there might not be agents, or actions, at all!

Surely we know (at least roughly) how to distinguish agents from non-agents, and actions from non-actions, and would continue to be able to do so even if the physicists at Caltech told us that causal determinism was true! But as I have already tried to argue, one need not accept this claim that the question whether causal determinism is true is a question we must leave to be settled by the physicists. We might instead suppose that a proper understanding of what actions are already falsifies the doctrine of causal determinism. And that prospect is only bewildering to someone who has already embraced fundamentalism. With causal determinism recognised for the exceedingly heavy metaphysical commitment I believe it to be, it should no longer seem extraordinary that we might hope to refute it by means of careful reflection on some of our *other*, and arguably less negotiable, metaphysical commitments.

The second important factor which has conspired to help cover up the possibility of an incompatibilism based solely on a metaphysical view of agency as the exercise of a two-way power, is the part which has been played in the debates surrounding moral responsibility and alternative possibilities by a range of real and imaginary addicts, neurotics, compulsives, phobics, etc. – whose problematic behaviours are thought of potentially excusable on the grounds that the agents in question 'could not have done otherwise'. It is held, quite reasonably, though, that it is most implausible to suppose that none of this problematic behaviour counts as a manifestation of the phenomenon of action. Surely, then, it is inferred, we must accept that there can be causally determined actions, and having accepted that, ought to turn swiftly to the important business of explaining what might be the necessary and sufficient conditions of those distinctively *free* actions for which we may truly be held

morally responsible. But is it really right to suppose that these cases show that there can be causally determined *actions*? Perhaps it might be causally determined that an alcoholic, say, might not be able to resist drinking (within a certain time period) the remains of a bottle of whisky left out in plain view in an easily accessible place. But the *particulars* of the action, or actions, involved will still be up to the alcoholic – where exactly to drink the whisky, and how quickly, and whether from the bottle or from a glass, and beginning right now or slightly later, diluted with water or not at all, etc., – and so surely we might insist that the action itself, considered as a particular event, was *not* a causally determined occurrence, but rather a genuine exercise of the distinctive two-way power of agency, as all true actions must be. That the occurrence of a certain behaviour-*type* has absolutely no chance of not occurring within a certain time-frame under certain conditions, does not imply that the token action which constitutes its actual occurrence was a causally determined event. Only where there are real reasons to insist that *none* of the particulars of some supposed 'action' was up to the agent – that he had no control over where, when, how or whether that action occurred – is it truly clear that it is wrong to suppose that what occurred really was an action – rather than merely an event in whose occurrence the agent played no meaningful role at all. Of course, if one is looking for a power of refrainment on the basis of which one might consider *excusing* such an agent for what he has done, the power to refrain from undertaking a given *particular* action is going to be far too flimsy. But that (as I have already argued) is not the only way in which a power of refrainment might figure in an argument for the view that causal determinism is inconsistent with moral responsibility. If one supposes that the power of refrainment

is important rather as a necessary condition of agency itself, it need not *also* function as the basis on which moral responsibility is withheld in these difficult cases.

A third reason for supposing that there is nothing about the idea of action *per se* which presupposes indeterminism, I think, is the very natural thought that many non-human animals can act. But this will only seem to present problems for a view of agency as essentially involving the exercise of two-way powers if one supposes that non-human animals cannot exercise such powers. And why should one suppose this? It is no doubt true that appeal to such phenomena as instinct will have to play a large role in the explanation of a great deal of animal activity. But what an instinct explains is generally why a certain *type* of behaviour occurs. It does not follow that an animal's execution of the actions required to comply with instinctual necessities must be conceived of as a series of deterministically produced events. One might rather suppose that just as we humans are sometimes unable to refrain from *types* of activity which nevertheless constitute our actions, in virtue of the fact that we can settle such matters as when, how and where, precisely, those actions will occur, so those higher animals to which we might wish to accord the power of agency generally retain control over many of the specifics of the activities in which they are at the same time (in a general way) instinctively compelled to engage – and that it is in virtue of that control that we are inclined to interpret their movements as the results of actions in the first place. For example, a cow is doubtless not free to refrain from grazing in the lush field in which she has been placed. But it does not follow that she cannot determine in which part of the field, exactly, she will choose first to graze; or interrupt her activity to lie down for a while; or settle which side of her mouth she will use to

chew the next mouthful of grass. Of course, in some of these cases, there will likely be reasons which help to explain why she chooses as she does. But as with a human agent, we might want to resist the idea that the existence of such reasons amounts to the existence of a deterministic cause of the movements of the cow.

It seems to me, then, that none of these arguments for supposing that the existence of actions has to be regarded as compatible with the thesis of causal determinism is at all compelling. It is obvious how the assumption of fundamentalism figures in the first argument. But it plays a role, I should argue, also in the second and third. For it encourages us, in considering the causal provenance of the actions of these non-responsible agents – the addicted, phobic and compulsive on the one hand, and those non-human animals to which we might want to accord agency, on the other - to focus exclusively on the complex chains of internal, neurophysiological, electrical and chemical events, which culminate eventually in the bodily movements of these agents – and to forget the ways in which we *ordinarily* conceptualise such activity as the exercise of power on the part of the individual agent concerned. And in so far as these two pictures of a movement's causal antecedents might look to be irreconcileable with one another, moreover, fundamentalism encourages us to think that the second must give way to the first. But anti-fundamentalism challenges that idea. It proposes that we make room for the suggestion that causality comes in many shapes and sizes, that it operates at many levels, and that all kinds of different objects (not just microphysical ones) may be the bearers of distinctive causal powers. The possibility arises, therefore, that agency might just be one of these distinctive types of power, one with its own enormously important place in the messy patchwork of

disparate phenomena which constitutes causal reality – not one which has to be made to fit into a world conceived of merely as a vast quantity of homogeneous particles blindly obeying physical laws. And if that were so, it would not be obvious that agents (or their powers) would simply have to dissolve, on inspection, into the flux of small things.

It seems to me, then, that anti-fundamentalism, rather than semi-compatibilism, provides the best hope for someone wishing to save moral responsibility from the threat presented by physics. Nevertheless, the difference between Fischer's view and my own should not be over-estimated. In particular, I think we ought to agree with Fischer that the whole 'alternative sequence' in which Black intervenes and brings it about that Jones votes for Clinton in any case is entirely *irrelevant* to the question whether or not Jones is responsible for what he does in the actual scenario – what matters is indeed, as Fischer maintains, only what happens in what he calls the 'actual sequence'. Where I differ from Fischer is only in my insistence that the question what *does* happen in the actual sequence is not independent of the question whether the agent has any alternative possibilities – for unless she possesses the power to refrain from the particular action she undertakes, I maintain, the actual sequence cannot constitute the occurrence of an action. I want to end with some brief reflections on Fischer's own conception of what is required of an actual sequence if it is to amount to an action for which an agent might be morally responsible.

As I see it, the tensions implicit in Fischer's attempt to make agency altogether independent of alternative possibilities surface eventually in the difficulties

he faces in attempting to clarify the second component of his account of guidance control – the need for the reasons-responsive mechanism which generates the agent's action to be 'the agent's own'. This second component is required in the first place because, as Fischer puts it, "one might exhibit the right sort of reasons-responsiveness as a result (say) of clandestine, unconsented-to electronic stimulation of the brain (or hypnosis, brainwashing, and so forth)" (p.18). But it is extremely difficult for someone in Fischer's position to offer a satisfactory account of what it is for a mechanism to be 'the agent's own' which will solve the problem of manipulation. On the one hand, he is not allowed to make any irreducible reference in his account of the generation of morally responsible forms of action to such things as the *agent* or her distinctive *powers* – since he is seeking an account which is entirely compatible with causal determinism thought of as a view about the event-causal determination of everything that occurs by purely physical laws, and it is plausible to suppose that no unreduced agents or such peculiar things as two-way powers could figure in such an account. On the other hand, any mere chain of events and circumstances, whatever it contains, and however complex it may be, would seem to be potentially reproducible in all its complex entirety by a manipulative agent, so that whatever set-up is offered by Fischer as being allegedly sufficient for the exercise of guidance control would seem to be open to the objection that that very set-up might itself be brought into existence (and maintained) by the manipulative agent – in which case, once again, it might seem that the manipulated agent could not be truly morally responsible for actions resulting from the set-up. For instance, Fischer's own suggestion that the mechanism on which one acts becomes "one's own" in virtue of one's having certain

beliefs about one's own agency and its effects in the world, seems obviously vulnerable to the objection that such beliefs could perfectly well be induced, alongside the original manipulative activity, by the manipulative agent. In that case, it would seem to be possible that I should have all the beliefs in question, and yet be simply deluded. All my actions are in fact merely the result of decisions made by the manipulative agent, and yet I *take* it that I have effects in the world and that I am an apt target for the reactive attitudes. Surely, if there is a genuine worry about manipulation in the first place, adding in such extra beliefs as these is not going to be enough to solve it. A compatibilist would do better, in my view, simply to bite the bullet and insist that mere reasons-responsiveness of the relevant mechanism *is* sufficient for moral responsibility, however counterintuitive that may at first appear to be.

It appears to me, moreover, that this same basic difficulty is going to infect *any* view which fails to assign actions the sort of metaphysically exceedingly distinctive nature I have been insisting they must have. Any view which descends from the level of agents to the level of such things as mechanisms, processes and events is going to face the problem that any mechanism, process or event which occurs inside an agent can be set in train by someone, or something, which is not the agent. Only if one accepts that an action is *essentially* the exercise of a power by the agent whose action it is can this difficulty possibly be avoided.

I do not say that it is going to be easy to develop the wanted account of actions within an acceptably naturalistic framework. My guess is that we are going to need to work hard to make better sense than we have so far made of notions like downward

causation and emergence; to make full use of the intellectual resources that appear to be emerging within such fields as complexity theory and systems biology; and to make better links than we have so far forged between the topic of consciousness and that of free will. But neither do I see any reason for supposing that it is bound to be impossible – *provided* we are willing to reject the position that I have been calling fundamentalism. Actions are natural phenomena, and so are agents. Our first step in understanding how they are possible should be the rejection of the idea that mere physics is what exclusively governs the unfolding of the world from one moment to the next.

Helen Steward
16/11/06