Supervenient Freedom and the Free Will Deadlock

Elzein, Nadine

2017-10

Elzein , N & Pernu , T K 2017 , ' Supervenient Freedom and the Free Will Deadlock ' , Disputatio , vol. 9 , no. 45 , pp. 219-243 . https://doi.org/10.1515/disp-2017-0005

http://hdl.handle.net/10138/299182
https://doi.org/10.1515/disp-2017-0005

DE GRUYTER
OPEN

Supervenient Freedom and the Free Will Deadlock

Nadine Elzein
University of Oxford

Tuomas K. Pernu
King's College London and University of Helsinki

BIBLID [0873-626X (2017) 45; pp. 219–243]
DOI: 10.1515/disp-2017-0005

Abstract
Supervenient libertarianism maintains that indeterminism may exist at a supervening agency level, consistent with determinism at a subvening physical level. It seems as if this approach has the potential to break the longstanding deadlock in the free will debate, since it concedes to the traditional incompatibilist that agents can only do otherwise if they can do so in their actual circumstances, holding the past and the laws constant, while nonetheless arguing that this ability is compatible with physical determinism. However, we argue that supervenient libertarianism faces some serious problems, and that it fails to break us free from this deadlock within the free will debate.

Keywords
Compatibilism; determinism; incompatibilism; non-reductive physicalism; supervenient libertarianism.

1 Introduction

It is widely acknowledged that debates about free will and determinism have an unfortunate tendency to end in deadlock. Recently, however, a number of theorists have put forward views that seem to offer us a serious prospect of breaking the sort of deadlock that has plagued traditional debates.

The sorts of views we wish to examine share a number of distinctive features. Firstly, their proponents are usually willing to commit

Disputatio, Vol. IX, No. 45, October 2017
Received: 27/07/2017 Accepted: 02/11/2017

to a libertarian (specifically leeway-incompatibilist) account of free agency, whereby agents are regarded as free only on the condition that they are able to do otherwise, holding the past and the laws of nature constant; secondly, their proponents usually accept a deterministic account of the laws of physics (or are, at least, willing to concede the possibility of physical determinism). And finally, proponents of such views claim that deterministic laws at the *physical level* are consistent with the presence of indeterminism at the *agency level*. While they concede that agency-level determinism would indeed be a real threat to free will, they maintain that determinism at the physical level alone poses no such threat.

Adherents typically maintain that the level of agency supervenes on the physical level, and that such a supervenience relation leaves open the possibility of distinctive level-relative modalities. We will call this sort of view *supervenient libertarianism,* and its proponents *supervenience libertarians*.[1] This view has recently been given a systematic analysis and defence by List (2014), and our discussion will largely rely on that particular treatment. However, closely related views have been proposed by Kenny (1978), Taylor and Dennett (2002), Berofsky (2010; 2012), Roskies (2012), Backmann (2013), Ismael (2013; 2016) and List & Menzies (2017) (*cf.* Pernu 2017).

We will argue that this view, if successful, would be a very significant advancement within the free will debate. However, we will also present reasons to doubt that it succeeds. While the sort of argument we find in favour of this view generally relies heavily on the supervenience thesis, it also requires us to overlook a crucial implication of this thesis; namely, that modalities at the level of agency are constrained by modalities at the physical level. While there may be cogent arguments in favour of disregarding this implication, an

[1] List (unpublished) dubs this view 'compatibilist libertarianism'. However, we want to avoid this term for it unites two contradictory terms and it is easy to confuse with 'libertarian compatibilism', a different view proposed by Vihvelin (2000). The main claims that characterise the view that we are focusing on are (a) that indeterminism is necessary for free will, and (b) that supervenient, agency-level indeterminism is compatible with subvenient, physical-level determinism. The term 'supervenient libertarianism' reflects this idea as clearly as possible. (However, in figure 1 below we also use the term 'actualist compatibilism', which comes closer to the spirit of the term that List (unpublished) is using.)

explicit appeal to such arguments brings the supervenience libertarian's commitments very closely in line with those of traditional compatibilists. Moreover, it is precisely a dispute about these sorts of commitments that has led to a deadlock between traditional compatibilists and incompatibilists. Hence, we argue that, despite initial appearances, supervenient libertarianism does not really provide such a dramatic departure from the traditional free will deadlock.

2 The traditional free will deadlock

2.1 Freedom, determinism, and the analysis of 'able to do otherwise'

Prior to Frankfurt's (1969) discussion (which raised serious doubts about whether the ability to do otherwise is a necessary condition of moral responsibility) the free will debate has traditionally focused on the question of whether alternative possibilities can be reconciled with determinism. Compatibilists have traditionally supposed that they can, while incompatibilists have traditionally supposed that they cannot.

While compatibilists have sought to defend a broadly conditional analysis of the ability to do otherwise (Moore 1903; Ayer 1954; Smart 1961; Schlick; 1966; Lewis 1981; Berofsky 2002) or, more recently, a dispositional analysis (Fara 2008; Smith 1997, 2003; Vihvelin 2004, 2011, 2014), incompatibilists have typically rejected this kind of reading in favour of a non-conditional analysis (Campbell 1951; Chisholm 1964; Lehrer 1968; van Inwagen 1983, 2000, 2004, 2008; Kane 1999).

On conditional and dispositional analyses, an alternative course of action is understood to be possible so long as it could, or would, happen in some given set of non-actual circumstances. On a conditional analysis, statements like:

(1) She could have done otherwise.

Are analysed as meaning something along the lines of:

(1) If she had chosen to do otherwise, she would have done otherwise.

On a dispositional analysis, statement (1) is analysed as meaning something like:

(3) If she had been placed in different circumstances, she would have done otherwise.

In contrast, on a non-conditional analysis, (1) will only be true insofar as we are able to make a non-conditional assertion about the agent's abilities, such as:

(4) She could have done otherwise as things actually were, holding the past and the laws of nature constant.

Clearly, neither (2) nor (3) suffices to establish (4). Much of the dispute has focused on the question of whether we ought to adopt a strong, non-conditional reading of 'able to' claims, as specified in (4) or whether a weaker conditional or dispositional reading, along the lines of (2) or (3) ought to be regarded as sufficient to capture the true meaning of (1).

Some new terminology will help us to capture the important contrast here more sharply. Let us divide those theorists who offer rival analyses of the ability to do otherwise into *actualists* and *non-actualists* (see figure 1 below). The former group is committed to a non-conditional analysis; they maintain that all and only the facts that characterise the actual features and history of a given set of circumstances matter when it comes to assessing what is possible in those circumstances. In contrast, the latter group, which includes both proponents of conditional and dispositional analyses are committed to the claim that only some of the facts need to be held constant, and that we must also consider what might have happened in some non-actual circumstances in assessing what is possible in any given situation.

This becomes a particularly important sticking point for ascertaining whether or not the ability to do otherwise is compatible with determinism. It seems as if the ability to do otherwise will be compatible with determinism insofar as we have reason to adopt a non-actualist analysis of this ability, and that it will not be compatible with determinism insofar as we have reason to instead favour an actualist analysis.

Serious disagreements arise when it comes to determining which

of these analyses best characterise the sort of ability that matters in grounding moral responsibility, *i.e.* when it comes to establishing which sort of ability we would need to possess in order to justify our present desert-entailing practices; those concerned with blame, praise, punishment, and reward. We cannot, therefore, resolve the dispute about whether the ability to do otherwise is compatible with determinism if we restrict ourselves solely to the realm of metaphysics. We also need to address ethical questions about the basis of those practices which our favoured ability claims are supposed to support.

2.2 The goal-relativity of 'able to' claims

The problem is that compatibilists and incompatibilists often have very different conceptions of both what stands in need of justifying and what the relevant sort of justification entails. This is especially problematic within this dispute, since 'able to' claims are essentially *goal-relative*.[2]

Suppose that someone were to ask you whether or not you could kill another human being. It is not at all obvious that this question has a simple 'yes' or 'no' answer, irrespective of considerations about *why* she might be asking it. The space of possible reasons why someone might be interested is potentially very broad. Perhaps she is thinking of hiring you as an assassin, and she wants to know whether you would do a competent job; perhaps she is a detective trying to narrow down her list of murder suspects; perhaps she is your martial arts instructor, and she is trying to determine whether you have mastered a certain lethal technique; perhaps she is a psychiatrist, and she's trying to work out whether you pose a risk to public safety and ought to be locked up; perhaps she is your police chief boss, and she's wondering whether to blame you for your failure to kill a notorious criminal mastermind when you had the chance.

It is not at all obvious that there is any single sense of 'able to' that will yield the right answer irrespective of what she is trying to

[2] The idea that 'able to' claims vary, at least with context, has been explored by a number of philosophers. Kratzer's (1977) analysis has been particularly influential. For a recent discussion of the way 'can' claims vary with context, see Maier 2015. We claim that 'able to' claims vary in relevance depending on the goals or aims of those using them, as opposed to varying with mere conversational context.

achieve. If she is your martial arts instructor, she will no doubt be interested in a conditional ability—she will want to know whether you would succeed in killing someone were you (for whatever reason) to put in a reasonable effort, utilising the technique that she has shown you. In contrast, if she is a psychiatrist trying to determine whether you pose a threat to public safety, she will not merely be interested in whether you could kill someone if you tried to. She will also need to know whether there is any serious risk that you will try to, as things actually stand. Whether you would kill if you were placed in a warzone or something is going to be irrelevant. Whether it is an actualist or non-actualist ability that matters depends crucially on what we are trying to pin on the ability. But compatibilists and incompatibilists are often engaged in rather different theoretical pursuits in relation to free will (a point which has been explored by others, see in particular Vargas 2005; 2009, *cf.* Elzein 2013).

Typically, compatibilists take our ordinary practices as a starting point, and then engage in a process of reflective equilibrium, aiming not merely to describe our usual practices, but also to explain them in an illuminating way, identifying the principles that actually underpin those practices. This procedure may reveal new normative pressures, pushing us to re-evaluate aspects of our practice, and to refine our principles, aiming for overall consistency within our account. But such an approach generally takes for granted that our usual judgements concerning responsibility and our practices as a whole are not wildly misguided. Some theorists even suggest that we ought to adopt a theory of free will that is highly 'resilient' in that its truth must not be thought to hang on whether or not certain scientific facts obtain (*cf.* Fisher 2006; Speak 2008; Capes 2013).

In contrast, incompatibilists tend to take seriously the possibility of global scepticism about freedom and moral responsibility, and hence tend to be doubtful about whether the principles that typically do underpin our practices really ought to. In light of this sort of moral concern, we cannot take for granted the broad accuracy of our ordinary ways of thinking about freedom. The sorts of principles we intuitively appeal to within our desert-entailing practices will not be a legitimate starting point for any process of reflective equilibrium on the matter. If this is the goal, resilience will obviously not be regarded as a virtue, since positing this in advance will be question

begging: it entails from the start that we will not take any threats to the possibility of free will seriously.

From the compatibilist's own perspective, success is often measured by a theory's ability to make sense of the broadest range of intuitive responsibility judgements with reference to the simplest and most compelling underlying principles. But in light of the incompatibilist's typical goals, this will not be regarded as a sufficient benchmark of success. If the goal is to justify our practices *as a whole* in response to serious moral doubts, the compatibilist project may seem overly descriptive and lacking in justificatory bite. If determinism is taken to be an external threat to the moral validity of the whole practice of blaming and praising, then any account that presupposes the broad validity of those practices will appear to be viciously circular.

For this reason, incompatibilists are often puzzled by the compatibilist's focus on non-actual conditions. They are typically driven by doubts about whether it is fair to blame an agent if that agent could not have done otherwise given the way things *actually* were. The incompatibilist understandably struggles to see why we would spend our time focusing on *non-actual* scenarios, when these differ in crucial respects from the situation that actually occurred.

Typically, incompatibilists will concede that if an agent had chosen otherwise she would have done otherwise, or that if the agent had faced a different situation, she would have chosen differently. But they will point out that the agent did not choose otherwise and that she did not face a slightly different situation. Moreover, if determinism is true, the agent could not have chosen otherwise or been placed in alternative circumstances either. If this is the nature of your worry, then you are unlikely to see the relevance of analyses that draw on what might have happened in non-actual circumstances; such analyses do not appear to provide the agent with alternatives that were suitably within their grasp in the relevant (*i.e.* actual) circumstances.

Given the different philosophical goals that typically underscore the sorts of projects compatibilists and incompatibilists take themselves to be engaged in, it is unsurprising that neither side is able to insist on their own favoured reading of 'able to' claims without begging an important question against their opponents. In order to justify the claim that one's favoured reading is the right one, one must

presuppose a set of goals which frame the problem from the start in a way that one's opponent is likely to reject.

2.3 Relative and absolute victories

How, then, are we to make progress when we hit this sort of stalemate? It seems that there are, essentially, three possible strategies here. Firstly, we may, like Frankfurt (1969), seek to bypass the dispute completely, arguing that the broader dispute between compatibilists and incompatibilists can be settled entirely independently of whether the ability to do otherwise is compatible with determinism. Secondly, we could give up on any goal of actually breaking the deadlock, aiming only to justify our own favoured reading of 'able to do otherwise' relative to our own conceptual goals. In this case, we may achieve a *relative victory*, but not an absolute one. A relative victory involves justifying your claim about whether the ability to do otherwise is compatible with determinism within the framework of your own project, albeit acknowledging that it would beg the question to insist that your opponent must adopt the same framework. Such a victory can be had insofar as you can justify your position on your own terms.

Or, finally, and much more ambitiously, we can aim for an *absolute victory*; the sort that requires one side to establish a claim about whether the ability to do otherwise is compatible with determinism on their opponent's own terms. In this case, the dispute will be won even if we adopt our philosophical opponent's favoured analysis of 'able to do to otherwise'. In relation to this debate, that might be done in one of two ways:

(1) Establishing that even if we grant a non-actualist reading of 'able to', determinism rules out the ability to do otherwise.

(2) Establishing that even if we grant an actualist reading of 'able to', determinism does not rule out the ability to do otherwise.

While Frankfurt (1969) tried to bypass this dispute by arguing that alternative possibilities are irrelevant to moral responsibility (and Frankfurtian compatibilists and source incompatibilists have followed his lead), the supervenience libertarian attempts to break the

deadlock, as opposed to merely bypassing it. The strategy involves attempting to show that agency-level indeterminism is compatible with physical-level determinism, and hence that agents may have the ability to do otherwise as things actually stand, holding the whole of the past and all of the laws constant, even in light of physical determinism.

Relative victories do move the discussion forward, in some ways. They enable us to better understand the divides that underpin the free will dispute. But they do not break the deadlock. An absolute victory, in contrast, would do so. Relative victories are easily won in this dispute, whereas absolute victories are hard to come by. If supervenient libertarianism succeeds, it would establish an absolute victory in the compatibilist's favour, and this would be a very important, even game-changing, step within the free will debate. Given that this argument has the potential to make such a significant step within the debate, it is worth giving it serious attention. Figure 1 shows a breakdown of the debate, and the place that supervenient libertarianism apparently seeks to occupy.

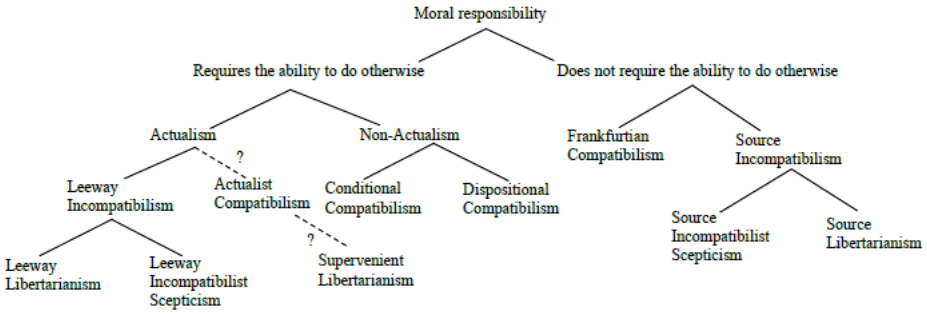

Figure 1. Taxonomy of the free will debate.

Many theorists, influenced by Frankfurt, bypass the dispute about how to analyse alternative possibilities by simply denying that alternatives are required for moral responsibility. This is represented on the right-hand side of the diagram. Although it has not entirely ended the dispute between compatibilists and incompatibilists, it has left those theorists engaged in a rather different dispute to the present one.

The present deadlock concerns the sort of disagreement we find on the left-hand side of the diagram. The problem to date is that

actualist analyses have tended to be motivated by conceptual goals that typically characterise the incompatibilist's main worries. In contrast, non-actualist analyses tend to speak only to the theoretical goals of the compatibilist.

Supervenience libertarians give the impression, however, that they are actualists: they maintain that an agent could have acted otherwise only if alternative possibilities were present for the agent, given the actual past history. Hence they adopt a reading of 'able to do otherwise' that seems to speak directly to the classic goals of incompatibilists, while at the same time establishing that this ability can be rendered compatible with determinism. We will offer a challenge to this view, and argue that it actually commits us to a form of non-actualist compatibilism very similar to the sort it seeks to reject. It is only in light of certain non-actualist assumptions, motivated by the broadly descriptive and explanatory goals that have typically driven compatibilists, that alternatives can be reconciled with physical determinism.

3 The supervenient libertarian view

3.1 Determinism

Let us begin with a formal definition of determinism (*cf.* List 2014), which may, for our purposes, be simplified as follows:

> **Determinism**
> A world w_1 is deterministic *iff* any possible world w_2 that shares the same laws and the same history prior to a given time t would share an identical history at all times subsequent to t.

And indeterminism is just the denial of this:

> **Indeterminism**
> A world w_1 is indeterministic *iff* there is some possible world w_2 that shares the same laws and history prior to a given time t, but whose history diverges from that of w_1 subsequent to t.

A world is deterministic, then, insofar as any world that shared its past history and its laws of nature would also share its future. But

determinism is normally understood to be a thesis about the way the laws of nature are at the physical level. Questions arise, then, about what bearing laws at this level have on the level of agency, and these push us to think more deeply about the relationship between these levels.

3.2 The physical level and the agency level

Non-reductive physicalists typically offer us a conception of the world according to which reality is carved into different levels, some of which have a more fundamental standing than others. Such a view has been popular in the philosophy of mind, where we have good reasons to suppose that mental states are dependent on brain states, but also have reason to suspect that they are not identical with brain states. We cannot understand mental phenomena and processes, and predict the behaviour of agents with reference to their neural states, and we cannot translate the sorts of language we use to describe mental states into neurobiological language. Nonetheless, we are aware that there is a very close connection between these levels. A very common way of understanding the relation between these levels appeals to the supervenience thesis. When phenomena at some level *A* supervene on phenomena at some other level *B*, this entails that there can be no change in the *A* phenomena without there being some corresponding change in the *B* phenomena.

This sort of relationship is less strict than identity, since the supervenience relation is asymmetric; it leaves room for changes in the subvening *B* phenomena with no change in the supervening *A* phenomena. In other words, the supervenience thesis allows for the multiple realisability of higher-level phenomena, whereby the same mental/higher-level phenomena could have had alternative physical/lower-level realisers. While phenomena at the base physical level fix all of the higher-level phenomena, the same higher-level phenomena may well be consistent with a range of different base-level underpinnings. Supervenient libertarianism utilises the supervenience thesis to establish level-relative modalities. This would allow for there to be alternative possibilities (analysed in actualist terms) at the level of agency, according to supervenient libertarianism, even if there are no alternative possibilities located at the base physical level.

3.3 Level-relative modalities and supervenient libertarianism

Supervenience libertarians thus claim that since multiple realisability allows for the possibility that two worlds could have identical histories at the level of agency, despite having different physical realisers, it also allows for the possibility that two worlds that share a history at the agency level may diverge from one another after a given point in time. It may be, the crucial claim is, that the different futures are possible in light of their different physical pasts, but since these pasts may realise the same agency-level phenomena, we can have indeterminism at the level of agency, despite the fact that the base physical level is deterministic. This is nicely illustrated with the following diagrams, provided by List (2014: 168):

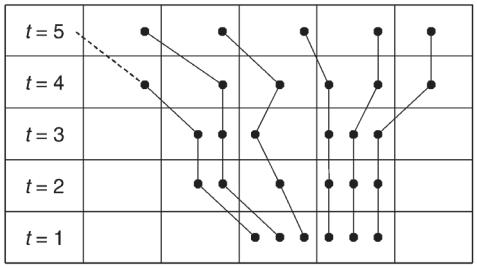

Figure 2.1: World histories at the physical level.

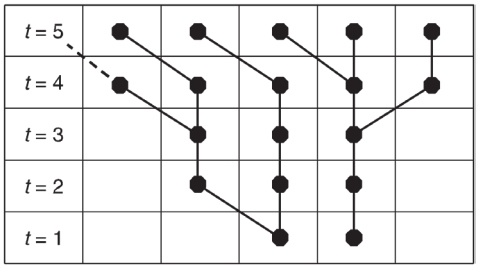

Figure 2.2: World histories at the agency level.

In figures 2.1 and 2.2 the vertical axis represents points in time, the squares represent mental or agency-level states, and the dots in the first picture represent different physical states that could realise these mental states. When two dots are located in the same square, these

are alternative physical realisers for the same agency-level states. The figure 2.1 represents the way things are at the base physical level. We can see that the different world histories include no branches; they are each deterministic.

The figure 2.2 gives us the world as it appears at the level of agency. And here, we do find branching. So it now seems that given that the same agency-level states could be realised by different subvening physical bases, it is possible that two worlds are exactly alike in terms of both history and laws at the agency level, but then diverge after a certain point in time. Hence agency-level indeterminism seems to be consistent with physical determinism.

Note, importantly, that on the supervenience libertarian view, these alternatives are not merely conditionally or dispositionally possible: an agent who is 'able to do otherwise', on this approach is not simply an agent who would have done otherwise *if* she had chosen differently, or been placed in different circumstances. Rather, at the level of agency, she could have done otherwise as things actually stand, holding the past and the laws of nature constant.

Thus, supervenience libertarianism seeks to show that an actualist reading of 'able to do otherwise' can be rendered consistent with determinism, and hence seeks to establish an absolute victory; it attempts to show that the ability to do otherwise is compatible with determinism, but to do so on the incompatibilist's own terms. If successful, this is a very important step. However, in the following section, we will present reasons to be doubtful about whether this approach really succeeds.

4 Some problems for the supervenient libertarian argument

There is a serious difficulty faced by supervenient libertarianism, which relates to a cluster of concerns about the justification for viewing modalities in a level-relative way and the coherency of its stance towards the supervenience thesis. We will argue that such concerns can be countered, but that doing so serves to narrow the gap between the supervenient libertarian view and traditional compatibilist accounts, significantly undermining the view's potential to seriously disrupt the deadlock that has traditionally plagued the free will dispute. We will proceed by outlining three closely related worries.

4.1 Taking supervenience seriously

Firstly, there seems to be an inherent tension within the supervenient libertarian argument, when it comes to the stance it must take towards the supervenience thesis. The conclusion of the argument asserts that modalities at the base physical level are broadly irrelevant with respect to those at the level of agency. But in order to generate level-relative modalities, the argument relies crucially on the supervenience and multiple realisability theses. The supervenience thesis, however, is precisely a claim about the way in which facts at the base physical level restrict what is possible at supervening levels: it says that once we have fixed the facts at the base-level, all of the higher-level facts follow from this.

We might think in terms of two distinct sorts of determinism operating at once. On the one hand, we have what we might term *horizontal determinism*, whereby, at the base physical level, future events are deterministically fixed by past events. On the other hand, we have something that we might term *vertical determinism*, whereby higher-level phenomena are fixed, by constitution or realisation, by lower-level phenomena. But there is a worry that the argument only really succeeds if we ignore the vertical part of this picture.

An example, borrowed from the debate about whether mental content is best understood to be broad or narrow, may help to illustrate why we suppose that there is a problem here. In using this example, we do not mean to suggest that the way in which we settle that dispute has any important bearing on this one, but the example helps to highlight something about the way in which modalities at more fundamental levels may bear on what is possible at supervening levels.

The example is that of jade, a material that was once thought to be a single substance, but is now understood to be two different substances, jadeite and nephrite. Chemical analysis has revealed that the microstructures of these two substances are actually quite different from one another, although they appear very similar in terms of their macro-qualities.

In reality, there are some clear differences between nephrite and jadeite on a macro-level, in terms of such things as strength, texture, and colour. For the purposes of this example, however, we ask you

to imagine that these substances are perfectly indistinguishable at the level of agency. Essentially, in this scenario, jadeite and nephrite would be alternative physical realisers for the macro-qualities that cause the same agency-level experiences involving jade.

Let us stipulate the following:

(i) Let w_1 and w_2 be two worlds that share the same laws of nature—these laws are completely deterministic at the physical level.

(ii) Let w_1 and w_2 have histories that are identical at the level of agency up until t.

(iii) At a microphysical level, however, let w_1 and w_2 differ in just one respect: all of the jade in w_1 is jadeite, whereas all of the jade in w_2 is nephrite.

(iv) In both worlds, Fatima conducts the first analysis of the microphysical structure of jade at t.

(v) After t, let the worlds diverge at the level of agency; in w_1 Fatima's analysis reveals a jadeite structure, and in w_2, it reveals a nephrite structure. Fatima's mental states and actions (*e.g.* what she writes in her notebook) subsequently also diverge.

Note, that this scenario ought to involve indeterminism at the level of agency, according to supervenient libertarianism. These worlds are indistinguishable prior to t at the level of agency; they share an exact history *prior* to t, and yet they have futures that diverge from one another *after* t.

But once we have established that w_1's future is possible for w_1 only because the stuff in w_1 is really jadeite and that the future of w_2 is only possible for w_2 because the stuff in w_2 is nephrite, it becomes hard to see why we ought to regard w_2's future as a genuinely possible continuation of w_1 and *vice-versa*.

Consider what is going on just prior to Fatima's analysis. She is in her laboratory with a piece of jade in front of her on the table, she sets up her apparatus competently, she understands exactly how to perform the relevant experiments correctly, and she knows that she will, in any case, uncover the true nature of jade when she conducts

her analysis. If the stuff sitting in front of her is in fact nephrite, it hardly seems possible in any meaningful sense that her analysis will truly uncover a jadeite structure. And if the stuff in front of her is actually jadeite, it seems similarly impossible that her analysis will truly uncover a nephrite structure.

Given that the possibilities are entirely fixed by the way things are at the level of microphysics, it seems very strange to assert that we can just ignore the bearing of microphysics when we come to assess which futures count as genuine possibilities for each of these worlds, given their laws and their pasts.

If this is dubious in the jade example, it seems no less suspicious in the case of agency more broadly. At this point, we may have doubts about whether we are entitled to describe the sorts of alternatives the supervenience libertarians identify as ones that are possible as things actually are, holding the past and the laws of nature constant. Rather, it now seems that what we will want to say is that *if* things had been different at the base physical level, then an alternative future *would* have been possible at the level of agency. But given the way things actually are at the physical level, we may have reason to doubt that this is a genuine possibility at all. Once we begin to see the view as really putting forward another variation of the non-actualist analysis, it becomes harder to see why it is supposed to have any greater appeal from an incompatibilist perspective than the traditional non-actualist analyses.

4. 2 Justifying the level-relative focus when comparing world histories

This first concern serves to highlight a second: we need a good justification for adopting a narrow, level-restricted focus in the first place, before we can even use this line of argument to generate these independent level-relative modalities.

That is because if we are to succeed in delineating different possible futures by levels, we need some way in which to justify carving the world up by levels, or else we will not be able to draw the relevant comparisons between possible world histories. In order to assert that two worlds share an identical history up until t, after which they diverge, we need some justification for focusing only on the higher-level, and ignoring its lower-level physical underpinnings in

our comparison. The question we must ask, then, is *why* we are entitled to carve the world into levels in this way, and to restrict our attention to the higher-level when we are comparing world histories.

It is worth noting that there are many potential ways in which we could carve the world up, and carving by supervening levels is merely one of them. In theory, any restricted focus, no matter how arbitrary, would generate special modalities relative to the portions of the world that we have restricted our attention to. But we would need a good reason for supposing that the restricted focus that we have chosen to adopt is important in some way. Clearly not all such restrictions will be philosophically illuminating. *E.g.* perhaps we could restrict our focus just to those parts of the world that don't contain hedgehogs, or just to the parts of the world that are within one metre of a spoon. Then we would generate rather different modalities—ones that are non-hedgehog-relative or close-proximity-to-a-spoon-relative, as opposed to level-relative.

Obviously, these are frivolous examples. But the point is that we would need some good reason to restrict our focus in this way before we embark on any attempt to generate these modalities. In relation to these examples, we don't think that such a restricted focus tells us anything deep about the independence of possibilities relative to these things. So we need to ask what's so special about the level-restricted focus we must adopt in this instance.

Why are we are entitled to ignore the microphysical state of the world in thinking about modalities, while we would not be entitled to ignore the hedgehog-containing parts of the world, or the parts that aren't near spoons? It does not follow simply from the fact that we can generate independent modalities by adopting this narrow field of focus that we ought to take such modalities seriously. However, some good reasons for adopting this narrow field of focus in relation to levels have actually been outlined. There are two basic lines of justification on offer.

Firstly, it is pointed out that we cannot successfully explain and predict events at the level of agency with reference to the way that things are at the base physical level. If we want to understand human behaviour and make successful predictions about it, then we had better restrict our focus to the level of agency. According to List (2014), 'mental-state ascriptions are indispensable in our current

best scientific explanations of human agency. Explaining and predicting even basic human interactions at a physical or neuroscientific level seems completely infeasible, whereas an intentional approach as simple as folk psychology has little difficulty making sense of them' (p. 168).

Secondly, further support can be offered for restricting our focus to agency-level phenomena by drawing on the fact that we are unable to translate the language that we use to describe agency into the language of physics, and cannot map the sorts of laws that govern human behaviour onto the laws that operate at the fundamental physical level. Our usual way of using language to describe the world seems to suggest that we restrict our talk about agency-level phenomena to the sorts of considerations that can be identified from within the agency-level perspective.

List (2014: 170) draws a fairly explicit parallel with Kratzer's (1977: 342–343) critique of one philosopher's suggestion (or 'misunderstanding', as she puts it), that it is meaningless for a judge to ask whether a particular murderer could have done otherwise, given that no one could have done otherwise if the entire state of the universe is taken into account. In fact, what the judge *meant* in the context, was to ask whether there were any features of particular relevant *aspects* of the situation (those located at the level of agency rather than, say, microphysics), that would have prevented the agent from acting otherwise. List (2014) echoes Kratzer (1977) in arguing that our views about freedom ought to track our ordinary-language semantics, capturing the way in which we typically use terms to describe phenomena relevant to freedom. The context in which judges ask about the abilities of defendants is situated entirely at the level of agency, and the language of abilities only really picks out restrictions that we can observe at the level of agency.

List (2014) further notes that, 'although neuroscientists have begun to identify a number of bridge laws that connect some specific cognitive phenomena with certain underlying patterns of brain activity, it is fair to say that a global reduction of psychology to physics is not in sight at this point. Indeed, if the central claims of nonreductive physicalism are correct, such a reduction is not feasible' (p. 171). He continues: 'Unless the reduction of psychology to physics led us to dispense with using psychological descriptions altogether—for

instance, if a grand unified theory of physics somehow subsumed all of psychology—we might still have to acknowledge an agent's ability to do otherwise *when employing the psychological level of description*' (p. 171).

This certainly does seem to provide some compelling reason to adopt a narrow field of focus, restricted by level, which differentiates this approach from other (more arbitrary) ways in which we could choose to carve up the world. But the justifications presuppose a particular sort of project and a particular set of goals relative to which adopting this narrow field of focus makes sense. And this brings us to another concern. We must now turn our attention to the sorts of goals that such justifications speak to, since goals, as we have already seen, have an important bearing for our understanding of ability claims and their significance.

4.3 The goal of moral justification *vs* explanatory and descriptive goals

The goals in relation to which the level-restricted focus is justified speak very much to concerns about how we describe, explain, and predict behaviour. These are interesting and important issues, which, in contrast to the earlier examples, can hardly be regarded as arbitrary. But the goals in question are broadly descriptive and explanatory, and this might call into question the relevance of these points when it comes to settling the dispute in a way that is supposed to speak to the traditional concerns of the incompatibilist—a crucial point if we are aiming for an absolute, as opposed to merely relative, victory.

These goals are quite far removed from the sorts of *moral* worries about the justification of our practices that incompatibilists are typically concerned with. In fact, this seems no less far removed from that project than previous compatibilist accounts, drawing on conditional and dispositional analyses of 'able to' claims. The explicit parallel with Kratzer's account makes this clear. Note that the philosopher accused by Kratzer of misunderstanding the judge may well have been concerned with more than simply interpreting what the judge *meant* to say. Perhaps his concern is that whatever the judge means, his sentencing will only be *justifiable* if some stronger sense of 'able to do otherwise' is established. Of course, meanings may vary

with context, but whether a course of action is morally justifiable is something that does not obviously shift with context.

This becomes a serious problem if we are hoping to break the free will deadlock by providing analyses on the incompatibilist's own terms. The argument looks like such an important step precisely because it reaches out to common ground. But at this point, the common ground starts to look illusory. The justifications here appeal to a particular set of aims, typically associated with the special sciences. In the special sciences, we are essentially trying to formulate predictively successful laws and intelligible explanations.

But intelligible explanations are the sort of thing that we might expect to be tethered to our limited and level-restricted perspectives as human beings. When we say that we cannot describe and predict agency-level phenomena solely with reference to phenomena at more fundamental levels, we might ask exactly who is to be included within the scope of 'we'? Present human beings? Supercomputers? Aliens? Lapace's demon? The average domestic cat? Future generations? What we can successfully describe and predict (as human beings with our limited perspective and the present state of technology) is an entirely contingent matter.

It is not at all obvious, however, in light of the moral concerns that underpin traditional incompatibilist worries, that the traditional incompatibilist is able to tolerate such contingency, given the nature of her goals. If we are assessing somebody's abilities with the aim of pinning some morally significant conclusions on the matter—for instance, with the aim of justifying some sort of retributive punishment—it seems quite important that our assessments are *not* restricted by our own contingent and limited perspective in this way. And if this is what we are morally concerned about in the first place, then the sorts of justification on offer are not obviously going to speak to those concerns.

To illustrate the point, consider the way in which shifts in technology might bring about shifts in our ability to successfully describe, explain, or predict human behaviour with reference to what is happening at more fundamental levels. If we invent better scanners and supercomputers, we might, in theory, be able to explain human behaviour with reference to more fundamental facts. But does the invention of a supercomputer have any important effect on whether

agents in general are really able to do otherwise? Might the invention of better scanners change whether or not an individual agent really *deserves* the death sentence, say?

For moral purposes it seems quite important not to reduce questions about what agents can do to questions about what we happen to be able to explain and predict. This, however, appears to be precisely what supervenient libertarianism does. We can only establish that agents are able to do otherwise consistent with physical determinism, by establishing level-relative modalities. We can only establish level-relative modalities, if we are justified in disregarding information about what is happening at the physical level when we assess what's possible at the level of agency. We can only justify disregarding this information by offering good reasons for focusing exclusively on the higher-level. But the justification we are offered for this appears to rest on rather contingent matters of fact—facts about what *we* happen to be able to describe and predict.

If you are sceptical about the idea that whether a given individual deserves praise or blame for a given act is the sort of thing that could depend on such contingencies of epistemology, then we ought to be sceptical about this argument. The problem once again comes down to conflicting goals, and begins to closely parallel the more traditional debate between compatibilists and incompatibilists. On the one hand, if our goals are roughly descriptive and explanatory, we may well expect them to shift with contingent shifts in what we are able to describe and explain with present methods. If, on the other hand, the goal is justificatory, and if we are willing to grant that there are serious doubts about moral responsibility in the first place, then we will not be willing to accept that such normative pressures can fundamentally change with shifts in technology or other contingent circumstances.

Moreover, once we have begun to cast doubt on some of these steps, we will similarly be doubtful about whether the supervenience libertarian really has analysed the ability to do otherwise in terms relevant for those who traditionally embrace an actualist analysis. Rather, it now seems merely that if things had been a bit different, agents would have done otherwise. At this stage, the relevance of the position to the traditional incompatibilist's concerns becomes far less obvious than it initially seemed. If we are genuinely interested

in what's outright possible, holding the past and the laws of nature constant, it seems that we might have good reason to suppose that the base physical level gets the final say here.

To say this, however, is not to say that the fundamental physical level has the final say irrespective of our theoretical goals. For some endeavours we do have good reason to carve the world up by levels, and to delineate modalities in level-relative ways. The dubious step comes, we think, when we move from the sorts of justification we have for taking this approach in one area of scientific or theoretical enquiry to conclusions about a separate area of enquiry, where possibilities are being assessed in pursuit of completely different goals.

This case provides a particularly stark example, since the argument tries to derive a *moral* justification for certain practices with reference to rather *pragmatic* considerations about language and predictive success, which are benchmarks for success only relative to a very different sort of enquiry, one aimed at goals that are solely explanatory and descriptive.

5 Conclusion

On the face of it, supervenient libertarianism looks like a promising strategy for breaking the traditional deadlock in the free will debate. However, on a closer inspection, it is not obvious that this new approach moves us so far from the traditional free will debate at all.

In particular, we have suggested that the opposing sides in the traditional free will debate focus on different sorts of ability, since compatibilists and incompatibilists are typically moved by different projects and goals, which render different readings of 'able to' claims relevant. In particular, the compatibilist is often trying to capture the principles that do in fact underlie our usual practices, where incompatibilists, in contrast, take seriously the prospect that our usual practices may be fundamentally misguided, rendering any account that takes the validity of those practices for granted question begging.

The challenge for supervenient libertarianism, however, is to explain why we are entitled to focus exclusively on the level of agency when answering questions about agency-level possibilities, ignoring the way in which facts about the physical world may have a bearing on what is and is not possible at the level of agency. It is when we

examine these justifications, that we start to see precisely the same sorts of assumptions which have always led compatibilists to define modalities in non-actualist terms. This approach is justified in relation to goals that the incompatibilist typically regards as irrelevant, and hence we find that we do not succeed in escaping the deadlock after all.

Does this mean that supervenient libertarianism does not move the debate forward at all? Our view is certainly not that the account does not contribute anything valuable to the debate. It helps to explain, on the compatibilist's own terms, what the relevant focus is for optimally serving the compatibilist's goals. What it does not do, however, is speak very loudly to the typical goals of incompatibilists. Hence supervenient libertarianism does not succeed in breaking the deadlock between compatibilists and incompatibilists in any meaningful way.[3]

Dr Nadine Elzein
University College, University of Oxford
Oxford, United Kingdom
nadine.elzein@univ.ox.ac.uk

Dr Tuomas K. Pernu
Department of Philosophy, King's College London
London, United Kingdom
Division of Physiology and Neuroscience
Department of Biosciences, University of Helsinki
Helsinki, Finland
tuomas.pernu@kcl.ac.uk

[3] We are grateful to Dr Maria Alvarez, Professor Bill Brewer, Professor Josep Corbi, Mr Taylor Cyr, Dr Julien Dutant, Mr Matthew Hart, Professor Ferenc Huoranszki, Dr Alexander Kaiserman, Dr Benjamin Matheson, Dr Eliot Michaelson, Professor Carlos Moya, Dr Pablo Rychter, Professor Carolina Sartorio, Dr Robyn Repko Waller, Dr John Wright, and the attendees of the King's College London Department of Philosophy staff seminar and the Il Blasco Disputatio conference in Valencia, 30th September, 2017, where earlier versions of this paper were presented, for valuable comments and discussions. Dr Pernu's work has been financially supported by the Kone Foundation and the Finnish Academy of Science and Letters.

References

Ayer, Alfred J. 1954. Freedom and necessity. In his *Philosophical Essays*, New York: St Martin's Press.

Backmann, Marius. 2013. *Humean Libertarianism: Outline of a Revisionist Account of the Joint Problem of Free Will, Determinism and Laws of Nature*. Heusenstamm: Ontos Verlag.

Berofsky, Bernard. 2002. Ifs, cans, and free will: the issues. In *The Oxford Handbook of Free Will*, ed. by Robert Kane. Oxford: Oxford University Press.

Berofsky, Bernard. 2010. Free will and the mind-body problem. *Australasian Journal of Philosophy* 88: 1–19.

Berofsky, Bernard. 2012. *Nature's Challenge to Free Will*. Oxford: Oxford University Press.

Campbell. Charles A. 1951. Is 'freewill' a pseudo-problem? *Mind* 60: 441–465.

Capes, Justin. 2013. Mitigating soft compatibilism. *Philosophy and Phenomenological Research* 87: 640–663.

Chisholm, Roderick. 1964. Human freedom and the self. In *Free Will*, second edition, ed. By Gary Watson. Oxford: Oxford University Press.

Elzein, Nadine. 2013. Basic desert, conceptual revision, and moral justification. *Philosophical Explorations* 16(2): 212–25.

Fara, Michael. 2008. Masked abilities and compatibilism. *Mind* 117: 843–865.

Fischer, John Martin. 2006. *My way: Essays on Moral Responsibility*. Oxford: Oxford University Press.

Frankfurt, Harry. 1969. Alternate possibilities and moral responsibility. *The Journal of Philosophy* 66: 829–839.

Ismael, Jenann. 2013. Causation, free will, and naturalism. In *Scientific Metaphysics*, ed. by Don Ross and James Ladyman. Oxford: Oxford University Press.

Ismael, Jenann. 2016. *How Physics Makes Us Free*. New York: Oxford University Press.

Kane, Robert. 1999. Responsibility, luck, and chance: reflections on free will and indeterminism. *The Journal of Philosophy* 96: 217–40.

Kenny, Anthony. 1978. *Freewill and Responsibility*. London: Routledge and Kegan Paul.

Kratzer, Angelika. 1977. What 'must' and 'can' must and can mean. *Linguistics and Philosophy* 1: 337–355.

Lehrer, Keith. 1968. Cans without ifs. *Analysis* 29: 29–32.

Lewis, David. 1981. Are we free to break the laws? *Theoria* 47: 113–121.

List, Christian. 2014. Free will, determinism, and the possibility of doing otherwise. *Noûs* 48: 156–178.

List, Christian. Unpublished. What's wrong with the consequence argument? In *Defence of Compatibilist Libertarianism*. Presently unpublished manuscript (accessible at url: http://philsci-archive.pitt.edu/11690/1/ConsequenceArgument.pdf)

List, Christian and Menzies, Peter. 2017. My brain made me do it: the exclusion

argument against free will, and what's wrong with it. In *Making a Difference*, ed by Helen Beebee, Christopher Hitchcock and Huw Price. Oxford: Oxford University Press.

Maier, John. 2015. The agentive modalities. *Philosophy and Phenomenological Research* 90: 113–134.

Moore, George E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Pernu, Tuomas. K. 2017. Can physics make us free? *Frontiers in Physics* 5: 64.

Roskies, Adina L. 2012. Don't panic: self-authorship without obscure metaphysics. *Philosophical Perspectives* 26: 233–342.

Schlick, Moritz. 1966. When is a man responsible? In *Free Will and Determinism,* ed. by Bernard Berofsky. New York: Harper and Row.

Smart, John J. C. 1963. Free will, praise and blame. *Mind* 70: 291–306.

Smith, Michael. 1997. A theory of freedom and responsibility. Reprinted in his *Ethics and the A Priori.* Cambridge: Cambridge University Press, 2004.

Smith, Michael. 2003. Rational capacities, or: how to distinguish recklessness, weakness, and compulsion. Reprinted in his *Ethics and the A Priori.* Cambridge: Cambridge University Press, 2004.

Speak, Daniel. 2008. Guest editor's introduction: leading the way. *Journal of Ethics* 12: 23–128.

Taylor, Christopher, and Dennett, Daniel. 2002. Who's afraid of determinism? Rethinking causes and possibilities. In *The Oxford Handbook of Free Will*, ed. by Robert Kane. Oxford: Oxford University Press.

van Inwagen, Peter. 1983. *An Essay on Free Will.* Oxford: Oxford University Press.

van Inwagen, Peter. 2000. Free will remains a mystery. *Philosophical Perspectives* 14: 1–20.

van Inwagen, Peter. 2004. Freedom to break the laws. *Midwest Studies in Philosophy* 28: 336–350.

van Inwagen, Peter. 2008. How to think about the problem of free will. *The Journal of Ethics* 12: 327–341.

Vargas, Manuel. 2005. The revisionist's guide to responsibility. *Philosophical Studies* 125: 399–429.

Vargas, Manuel. 2009. Revisionism about free will: a statement and defense. *Philosophical Studies* 144: 45–62.

Vihvelin, Kadri. 2000. Libertarian compatibilism. *Philosophical Perspectives* 14: 139–166.

Vihvelin, Kadri. 2004. Free will demystified: a dispositional account. *Philosophical Topics* 32: 427–450.

Vihvelin, Kadri. 2011. How to think about the free will/determinism problem. In *Carving Nature at its Joints,* ed. by Joseph K. Campbell and Michael O'Rourke. Cambridge, MA: MIT Press.

Vihvelin, Kadri. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter.* New York, NY: Oxford University Press.