Probabilistic supervenience and agential possibilities[1]

Alexander Gebharter * Maria Sekatskaya

Abstract: Compatibilist libertarianism proposes a new solution to the problem of an apparent incompatibility of free will and determinism. It drives a wedge between ontological levels and claims that free will is possible as a higher-level phenomenon even if the fundamental physical level is governed by determinism. After highlighting an inconsistency in the current version of compatibilist libertarianism, we discuss how one of its essential metaphysical assumptions (in particular: supervenience) can be modified in order to avoid this problem. Finally, we discuss the pros and cons of pushing the position to the limits in this way.

Keywords: free will, determinism, supervenience, multiple realisability, non-reductive physicalism

1 Introduction

We encounter the assumption that we possess free will in one form or another in our everyday lives, but also when we look at how we behave as a society or how we think about others' actions in legal or moral contexts. We value the freedom to choose between political views and religions, to choose our job according to our interest, ambition, or convenience, and the freedom of speech. We hold others responsible for their actions if they were carried out freely, and tend to forgive them for the consequences of their actions if they did not result from such a free

[1] This is a draft paper. Please do not cite or quote. The final publication is published under the following bibliographical data: Gebharter, A. & Sekatskaya, M. (2025). Probabilistic supervenience and agential possibilities. Philosophical Explorations. Advance online publication. dot:10.1080/13869795.2025.2502373

choice. Almost everyone agrees that we possess free will in one way or another.[2] One of the big questions in the free will debate is what kind of free will we actually have. A whole plethora of positions is available in the literature. From an intuitive point of view it seems that the kind of free will which grants an agent the ability to do otherwise is especially attractive. According to this understanding of free will, an agent being free to choose a specific job means that she could have chosen a different one. It was up to her that she chose this specific job rather than a different one or being unemployed. An agent's free decisions make a difference for how the future unfolds. In this paper, we are interested in this particular type of free will.[3]

There are different ways to further distinguish the many positions to be subsumed under the overall account of free will as an agent's ability to do otherwise. For our endeavour, especially the distinction between actualist and non-actualist accounts is useful (cf. Elzein & Pernu, 2017). According to actualist accounts, an agent's having free will requires her to have the possibility of doing otherwise given the actual past (van Inwagen, 1983). Non-actualist accounts, on the other hand, are a bit less demanding. For them an agent has free will if she would have done otherwise in a close enough possible world in which the past might have been different than in the actual world to some extent (Campbell, 1997, 2005; Lehrer, 1990). Now there are two metaphysical possibilities: Either determinism is true or it is not. Determinism naturally poses a threat to actualist accounts. Determinism rules that given the past, there is only one possible way for the future to unfold. But then it seems that the actualist agent runs into trouble. Given the actual past, at no point in time does she have the possibility of doing otherwise. Her future actions are already fully determined by the past. This is why actualist accounts are typically held by incompatibilists about free will and determinism. Compatibilists, on the other hand, often hold non-actualist accounts. For these accounts, determinism is no problem. Though the past fully determines the future in each single possible world due to determinism, an agent could still

[2] Exceptions are free-will sceptics (Pereboom, 2001, 2014).

[3] This is the type of free will advocated by libertarians (Balaguer, 2010; Ekstrom, 2000; Ginet, 1990; Kane, 1996, 2019), classical or traditional compatibilists (Campbell, 1997, 2005; Lehrer, 1990), and supporters of dispositional accounts (Fara, 2008; Vihvelin, 2004, 2013).

have done otherwise. Had things been a bit different in the past, in particular, had some things about the agent been different, then also her future actions would have been different.

This paper is about compatibilist libertarianism. Compatibilist libertarianism was mainly developed by List (2014, 2019b) who proposed it as a position in the tradition of authors such as Dennett (2003), Kenny (1978), and Taylor and Dennett (2002). It is intended to provide a metaphysical foundation for reconciling actualist free will with determinism by distinguishing between the agential and the physical level and arguing that determinism at the latter is still compatible with choosing between different futures at the former. If this project succeeds, this would be excellent news for everyone favouring an actualist understanding of free will. In our view, actualist free will is the strongest kind of free will because it allows the agent to have the most control over the fate of the actual world. Arguably, it is also the most desirable to have.

Note List (2014, 2019b) is not entirely clear about what is required for the type of control compatibilist libertarianism is after. One obvious candidate would be causal control, which would require the agent to causally influence the future. We follow List in leaving the exact further requirements for control open in order to keep the arguments discussed as general as possible. What is important though is that control requires that there are at least two possible ways of how the future can unfold given that the past is fixed. Also note that we are not interested in the question whether actualist free will exists, but rather in whether the type of control compatibilist libertarianism requires is compatbile with determinism.

This paper is structured as follows. In section 2, we introduce compatibilist libertarianism and explain how it is supposed to reconcile actualist free will with determinism. In section 3, we discuss the collapse argument (Gebharter, 2020) which poses a serious problem for compatibilist libertarianism. We believe, however, that this is not the end of the story. In section 4, we argue that the incompatibility of actualist free will and determinism follows from one of the central assumptions of compatibilist libertarianism, namely, that agential states supervene on physical states. This means that there is no change in agential states without a change in physical states. We then tweak the supervenience assumption in such a way that compatibilist libertarianism can finally overcome the problem highlighted in section 3. In particular, we

propose replacing supervenience by a weaker probabilistic version of supervenience.[4] We are well aware that we push compatibilist libertarianism to the limits with this move and that this strategy does not come without costs. In section 5, we discuss the extent of freedom that tweaking compatibilist libertarianism in the suggested way can actually provide, highlight several possible problems for the modified position, and suggest ways to address these problems. We conclude in section 6.

2 Compatibilist libertarianism

Compatibilist libertarianism aims at rendering invalid one of the standard arguments against the compatibility of free will and determinism (which can be seen as a version of the consequence argument, cf. Hausmann, 2020; Kapitan, 2002; List, 2019a; van Inwagen, 1975, 1983). Determinism rules that the past together with the laws of nature fully determines the future. A simplified version of the argument that will do the job for this paper runs as follows:

P1: Free will requires the agent's ability to do otherwise (given the actual past).

P2: If determinism is true, then no agent can do otherwise (given the actual past).

C: If determinism is true, then no agent possesses free will.

Now the main idea underlying compatibilist libertarianism is to disambiguate an agent's ability to do otherwise in P1 and P2. There are two ways in which an agent can do otherwise: The agent can have the physical possibility or the agential possibility of doing otherwise. In List's (2014) words, "[t]he key idea is that although determinism implies that only one future sequence of events is *physically possible* given the current fully specified state of the world, the more coarsely defined state of an agent and his or her macroscopic environment can still be consistent with more than one such sequence, and thus different alternative actions can be *possible for the agent*" (p. 157).[5] Distinguishing agential from physical possibilities allows

[4] A similar move has recently been suggested in (Gebharter & Sekatskaya, 2024) as a novel strategy to block the causal exclusion argument from the mental causation debate.

[5] For further illustration of agential possibility in terms of branching of world histories, see the discussion of Figure 1 below.

compatibilist libertarianism to support the actualist notion of free will. For easier reference, let us give this specific understanding of free will a label:

(FW) An agent has free will if she has the agential possibility of doing otherwise (given the actual past).

A similar distinction can be made between determinism at the physical level and determinism at the agential level. Supporters of compatibilist libertarianism first and foremost interpret determinism as physical determinism. Let us introduce a label for this view as well:

(DT) Each physical state is fully determined by its physical past and the laws of nature.

Once one keeps the two levels and the corresponding versions of actualist free will and determinism apart, the argument above transforms into the following updated version:

P1*: Free will requires the agential possibility of doing otherwise (given the actual past).

P2*: If determinism is true, then no agent has the physical possibility of doing otherwise (given the actual past).

C: If determinism is true, then no agent possesses free will.

It is easy to see that the updated version of the incompatibility argument is invalid. To get conclusion C out of P1* and P2*, a further premise such as P3 is needed:

P3: Not having the physical possibility of doing otherwise (given the actual past) implies not having the agential possibility of doing otherwise (given the actual past).

Now compatibilist libertarianism is built on the basis of the observation that P3 is needed in addition to P1* and P2* to derive C. Thus, the overall success of the compatibilist libertarian's enterprise crucially hinges on whether the position is able to make convincing metaphysical assumptions according to which P3 comes out as false. In the remainder of this section, we will introduce, step-by-step, these assumptions and show how they are supposed to render P3 false.

First, let us come back to the distinction between the agential possibility of doing otherwise and the physical possibility of doing otherwise. Drawing this distinction does not yet guarantee that the two possibilities do not coincide. Suppose, for example, reductive physicalism were

true, meaning that non-physical properties can be reduced to fundamental physical properties. But then having the agential possibility of doing otherwise would be nothing over and above having the physical possibility of doing otherwise and premise P3 would come out as true. Because of this, compatibilist libertarianism needs to reject reductive physicalism. Since the position takes some major inspiration from modern science and the modern scientific picture of the world, it explicitly commits itself to non-reductive physicalism instead. Originally, non-reductive physicalism was intended as a position that allows (to some extent) for the autonomy of higher-level phenomena (cf. Kim, 2005). Non-reductive physicalism is committed to three characteristic assumptions:

(NR1) Higher-level properties are distinct from and not reducible to lower-level properties.

(NR2) Higher-level properties supervene on lower-level properties.

(NR3) Higher-level properties are relevant for explanation in the sciences.

Assuming that agential properties populate a higher level than physical properties, **(NR1)** implies that agential states are distinct from and not reducible to physical states. We understand states as instantiated properties. Furthermore, physical states are instantiated physical properties and agential states are instantiated agential properties. As common in standard versions of non-reductive physicalism, the concrete entities which instantiate higher- or lower-level properties are assumed to be physical. Thus, non-reductive physicalism is committed to monism about entities and dualism about properties.

The purpose of **(NR1)** for non-reductive physicalism is to drive an ontological wedge between the agential and the physical level that renders P3 contingent. With this assumption in place, P3 is not automatically true (like under the assumption of reductive physicalism). The second assumption, **(NR2)**, establishes a minimal connection between the agential and the physical level. In particular, it implies that there is no change at the agential level without a corresponding change at the physical level. Hence, it anchors all the phenomena investigated in the sciences in physics and, by doing so, forbids agential states which would be completely independent of physics. **(NR3)** implies that agential properties are sometimes relevant for

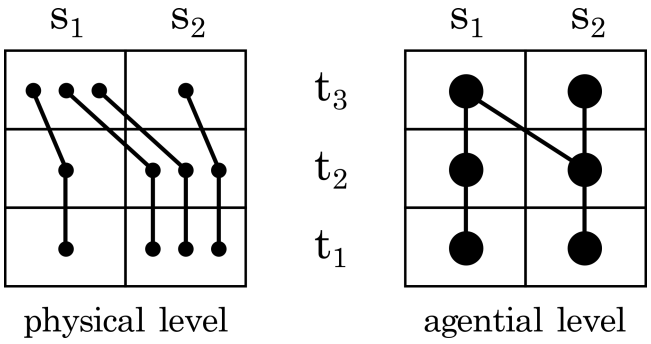

Figure 1

explanation in the sciences, which sounds plausible, but will play no further role in the remainder of this paper. Because of that, we will bracket **(NR3)** from now on.

Finally, one further metaphysical assumption, viz. multiple realisability, is essential for compatibilist libertarianism to block the incompatibilist argument:

(MR) Some agential states are multiply realisable by physical states.

This latest assumption guarantees that at least some agential states – for example, feeling pain – can be realised by different physical states – for example, different states of the brain. We will see which particular role it plays for compatibilist libertarianism shortly.

Next, we briefly summarise how the metaphysical assumptions introduced so far allow compatibilist libertarianism to block the incompatibilist argument. To this end, we introduce a simple (informal) model satisfying all the metaphysical assumptions made above. We then show that the actualist notion of free will **(FW)** can be satisfied too, which proves that actualist free will as a phenomenon restricted to the agential level is compatible with physical determinism. In other words, what the model will show is that given compatibilist libertarianism, P1* and P2* come out as true, but P3 as false, meaning that the incompatibilist argument sketched earlier is blocked and its conclusion C does not follow anymore.

The model is depicted in Figure 1. The left hand side shows the physical level and the right hand side the agential level. Accordingly, the small dots stand for physical states and the big dots for agential states. t_1 stands for a particular point in time with t_2 being later than t_1 and t_3 being later than t_2. For easier reference, the left and the right columns of each plane are labelled s_1 and s_2, respectively. The particular cells indicate how agential states can be realised at the physical

level. In particular, an agential state in a cell s_it_j can be realised by the physical states in the corresponding physical cell s_it_j. Vice versa, an agent being in a specific physical state in a cell s_it_j will also be in the agential state in the corresponding agential cell s_it_j. Finally, lines indicate the time evolution of the system at the physical and agential planes. We will also refer to them as world histories. How a system evolves at the agential level is determined by its evolution at the physical level. A system can evolve into an agential state s_it_j only if one of the physical states realising its initial agential state evolves into a physical state s_it_j as well.

With the model in place, let us now briefly run through the assumptions made by compatibilist libertarianism. Firstly, we can observe that **(NR1)** is satisfied: The ontological difference between the agential and the physical level is indicated by having two planes instead of one. More interestingly, also **(NR2)** is satisfied: Agential properties supervene on physical properties. At any time t_i, a change from one of the agential states would be accompanied by a change in physical states. Assume, for example, that the system is in the agential state in s_1t_1. Then it will also be in the physical state in s_1t_1. Had it been in the alternative agential state in s_2t_1, however, also its physical state would have been different. Its physical state would have been one of the three in s_2t_1 then. Finally, we can observe that also **(MR)** is satisfied: Some agential states are multiply realisable. In particular, these are the agential states in s_1t_3, s_2t_1, and s_2t_2.

Next, to show that physical determinism **(DT)** and actualist free will **(FW)** are compatible according to compatibilist libertarianism, we need to show that **(DT)** and **(FW)** both come out as true in the model. **(DT)** is true because for every physical state there is only one possibility for how the system's future can unfold at the physical plane, meaning that there is no branching of physical world histories towards the future. And finally, also **(FW)** is satisfied by the model: An agent who starts with the agential state in s_2t_1, for example, can end up either in the agential state in s_1t_3 or in the one in s_2t_3. There is branching of agential world histories towards the

future and, thus, the agent has the possibility to end up with different agential states given the actual agential past.[6]

So there you have it: Finally, it seems that we can have actualist free will (restricted to the agential level) as well as (physical) determinism.[7] The model renders P1* and P2* true, but P3 false: The fact that there is no possibility of doing otherwise (i.e., branching of world histories) at the physical level actually does not imply that there is also no such possibility at the agential level. Because of this, C does not follow and the incompatibilist argument is blocked. In the next section we discuss the collapse argument, which points to a blind spot in the argumentation that, once recognised, unmasks the argument for the compatibility of actualist free will and physical determinism outlined in this section all as smoke and mirrors.

3 The collapse argument

The collapse argument was put forward by Gebharter (2020). In its original version, it is formulated in terms of a probabilistic model. For our endeavour, however, we can ignore most of the technicalities and use the simpler terminology introduced in section 2 instead. According to the collapse argument, the main problem of the strategy to avoid the incompatibilist argument outlined in section 2 is that it relies on an inadequate identification. In particular, the argument showed that under the metaphysical assumptions made by compatibilist libertarianism, there can be branching at the agential level even if the physical level is deterministic. This branching at the agential level was then identified with an agent's ability to do otherwise. This identification, however, is misplaced. Branching is, upon closer examination, not enough for an agent's possibility of doing otherwise.

---

[6] There is a somewhat orthogonal argument to be found aiming at blocking the causal exclusion argument (Kim, 2005). Since it has no direct bearing on our endeavour in this paper, we content ourselves with pointing to some of the relevant literature: (Baumgartner, 2010, 2018; Blanchard & Hüttemann, 2024; Gebharter, 2017; Gebharter & Sekatskaya, 2024; Hoffmann-Kolss, 2021; Woodward, 2015, 2022).

[7] Note that this does not yet show that there *is* actualist free will. As mentioned before, we are only interested in the question of whether compatibilist libertarianism can render actualist free will compatible with determinism in this paper.

The collapse argument shows why this is so: Granted, there can be branching at the agential level according to compatibilist libertarianism. However, assume two agential trajectories branch out in different futures at a point in time t_j such that two different agential states seem possible for the agent at a later point in time t_k. Since we assume **(NR2)**, the actual agential state at t_j is realised by a particular physical state at t_j. Which physical state actually realises the agential state at t_j is, due to **(DT)**, fixed and fully determined by the actual world's physical state in the past, say at t_i. From **(DT)** it also follows that there is only one way the physical future can unfold from t_i onwards. At t_j, it will be in the physical state it actually is and at any later point in time, say t_k, it will be in another specific physical state. But now only one of the two agential states at t_k that seemed possible for the agent at t_j is actually accessible to her, because only one of them can be realised by the actual physical state at t_k. The latter is a direct consequence of **(NR2)**: Different agential states must have different physical realisers. But what this means is that the agent can in the end only access the agential state at t_k that is realised by the actual physical state at t_k into which the world will ultimately evolve due to **(DT)**. This shows that branching at the agential level does not mean that all the branches are actually accessible to the agent. In fact, the opposite is the case: Though different possible futures might seem possible for the agent, only one is actually accessible to her.

Let us further illustrate the argument by means of Figure 2: Assume the current actual agential state is the one at s_2t_2 within the circle and that the physical state realising it is the one on the physical plane in the circle. That this specific physical state realises the actual agential state is nothing that would be up to the agent by any means. The agent's current physical state at t_2 is, according to **(DT)**, fully determined by the physical past, i.e., by the earlier physical state represented by the black dot in s_2t_1. It also follows from **(DT)** that only one physical future is possible from the actual physical state at t_1 onwards. This physical future is the only one accessible to the agent. It is represented by the black dots and lines. It ultimately leads to the physical state in s_2t_3. All the grey lines on the physical plane are inaccessible, they are ruled out

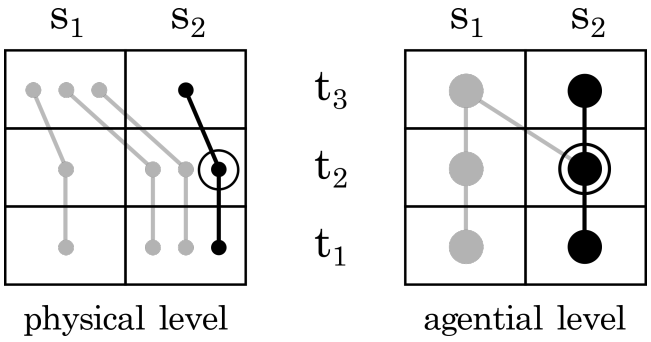

Figure 2

by the world's actual physical state at t_1 and **(DT)**. But this means the branching we see at the agential level at s_2t_2 is illusory. Though it seems that the agential state at s_2t_2 can develop into the one in s_1t_3 as well as the one in s_2t_3, the agential state in s_1t_3 is actually ruled out, because all of its physical realisers in the corresponding cell on the physical plane are ruled out by **(DT)** and **(NR2)** which do not allow for the actual physical state at s_2t_3 to be a realiser of more than one agential state. Hence, branching at the agential level does not mean that the agent actually has more than one possible course of action available to choose from. All the future possibilities seemingly available to her at the agential level collapse into a single one that is fully determined by the world's earlier physical state at t_1.

What the collapse argument shows is that P3, contrary to the compatibilist libertarians' initial assessment in section 2, actually comes out as true. The assessment in section 2 was defective because it conflated branching of agential trajectories with possible future courses of action available to the agent.[8] Based on this objection we must conclude that the incompatibilist argument outlined in section 2 still stands, regardless of the attempts to avoid it by subscribing to the metaphysical assumptions introduced by compatibilist libertarianism. In the next section,

[8] Our argument presupposes that only those future agential states are accessible for the agent whose physical realisers lie on a trajectory which can be traced back to the physical realisers of the agent's current agential state. This basic principle seems hard to doubt. Illustrated by means of a concrete example: The future decision to raise your left arm is accessible to you only if one of its physiological realisers can be traced back to your current brain state. Without this, the future decision to raise your left arm is not accessible to you on the agential plane because it cannot be realised on the physical plane given the current physical state realising your current agential state in the actual world.

we will push compatibilist libertarianism to the limit by modifying one of the metaphysical assumptions to which the position committed itself: the supervenience assumption **(NR2)**.

4 Compatibilist libertarianism pushed to the limits

In this section we propose a possible modification of compatibilist libertarianism that can avoid the collapse argument.[9] In our view, this is the only reasonable possibility to formulate a version of compatibilist libertarianism that in the end turns out to be an actualist account of free will.[10]

Let us begin by recalling the ingredients of compatibilist libertarianism: ontological non-identity and non-reducibility **(NR1)**, supervenience **(NR2)**, multiple realisability **(MR)**, physical determinism **(DT)**, and free will understood as the agential possibility of doing otherwise **(FW)**. Which one of these metaphysical assumptions can we modify? **(FW)** states the very meaning of free will that is essential to compatibilist libertarianism and, thus, needs to stay as it is. Since interpreting determinism as physical determinism is key to the compatibilist libertarian's strategy to render the incompatibility argument invalid (see section 2), we also need to keep **(DT)**. As we saw in section 2, **(MR)** is essential for branching at the agential level (at least when the system's physical state is not taken into account) and, thus, seems to play into the hands of compatibilist libertarianism rather than doing harm. But what about the core assumptions of non-reductive physicalism? **(NR1)** needs to stay for the reasons mentioned in section 2: It is essential to drive a wedge between the agential level and the physical level which is, in turn, crucial for any possible success in blocking the original incompatibilist argument. Finally, only **(NR2)** remains. This assumption was, as we saw in section 3, actually quite troublesome. The fact that supervenience fixes the agential state once its physical supervenience base is fixed,

---

[9] Menges (2021) recently proposed another possible strategy to avoid the consequence of the collapse argument. He suggests to shift the focus to an agent's ability to do otherwise from a moral perspective. We do not further discuss Menges' account in this paper since we are only interested in the metaphysical modal possibility of doing otherwise List (2014, 2019b) explicitly endorses.

[10] To avoid the collapse argument one can also interpret compatibilist libertarian as a classical compatibilist account (Sekatskaya & Gebharter, 2024). However, this move would render it a non-actualist rather than an actualist account of free will.

which it always is because of determinism, was one of the main culprits in the objection to the original version of compatibilist libertarianism. It made all the higher-level future paths of action seemingly available to the agent collapse into a single one. So let us have a closer look at **(NR2)**.

We can tweak the implications of **(NR2)** by revising the notion of supervenience figuring in **(NR2)**. Up to now, we understood supervenience in a strict sense: Any change in the supervening agential property is associated with a change in its physical supervenience base. This notion of supervenience can be further decomposed as follows:

(SUP) Agential properties $A_1, \ldots, A_m$ supervene on physical properties $P_1, \ldots, P_n$ if and only if

(i) any change in A_i-properties is associated with a change in P_j-properties, and

(ii) exact similarity in P_j-properties guarantees exact similarity in A_i-properties.

Both conditions are clearly satisfied by the original model in Figure 1: (i) is satisfied because changing the agential state (i.e., jumping from one dot to another in the same row at the agential plane) is associated with a change in physical states (i.e., jumping from one dot to another in the same row at the physical plane) and (ii) is satisfied because whenever the system is in a physical state (i.e., a dot in s_1 or s_2) it will be in its corresponding agential state (i.e., a dot in s_1 or s_2, respectively) as well.

Note that in the strict version of supervenience **(SUP)**, (ii) is actually the contraposition of (i), meaning that (i) and (ii) are logically equivalent. Nevertheless, for our endeavour it makes sense to keep conditions (i) and (ii) apart, as we will see shortly. To this end, we reformulate the strict version in probabilistic terms:[11]

(SUP)* Agential properties $A_1, \ldots, A_m$ supervene on physical properties $P_1, \ldots, P_n$ if and only if

[11] In the following, expressions such as "F" and "G" figuring in probability statements such as "$Pr(F \mid G)$" are to be interpreted as "property F is instantiated" and "property G is instantiated", respectively.

(i)* for every pair of non-identical properties A_i, A_j and every P_k: If $Pr(P_k|A_i) > 0$, then $Pr(P_k|A_j) = 0$, and

(ii)* for every P_k-property there is an A_i-property such that $Pr(A_i|P_k) = 1$.

The conditions (i)* and (ii)* of the strict probabilistic version of supervenience **(SUP)*** capture the same basic idea as their counterparts (i) and (ii) in **(SUP)**: (i)* guarantees that any change at the agential level is associated with a change at the physical level, while (ii)* says that once the physical state is fixed, the agential state is as well. In Figure 1, for example, (i)* is satisfied because whenever the agential state changes, the probabilities for physical states possibly realising the initial agential state change to 0. If we change, for example, the agential state from the one in s_1 to the one in s_2 at t_1, the probability of the physical state in s_1 at t_1 changes from 1 to 0. Vice versa, if we were to change the agential state in s_2 to the one in s_1 at t_1, the probability of each of the physical states in s_2 at t_1 would change from $r > 0$ to 0. And (ii)* is satisfied because each physical state in any of the cells renders the probability of its corresponding agential state 1. The physical state in s_1 at t_1, for example, renders the probability of the agential state in s_1 at t_1 equal 1 and each one of the physical states in s_2 at t_1 renders the probability of their corresponding agential state in s_2t_1 equal 1.

To avoid the collapse argument outlined in section 3, we need to relax condition (ii)* of **(SUP)***. The formulation of **(SUP)*** in terms of probabilities given above makes it clear how this could be done. In particular, we need to replace (ii)* by the following condition:

(ii)** for every P_k-property there is an A_i-property such that $Pr(A_i|P_k) \neq Pr(A_i)$.

This revised condition (ii)** does not require that each physical state fixes the agential state anymore. It only demands a change of probabilities for some agential state. In other words, it establishes a minimal connection between the agential and the physical level by granting a probabilistic effect of each instantiation of one of the physical properties on some agential property. We will see how this allows one to block the argument from section 3 shortly.

First, let us note that to arrive at a consistent notion of supervenience that finally allows for branching at the agential level, another tweak is required. If we relied on (i)* and (ii)**, the resulting notion of supervenience would either lead to the same problem as the original one or we would end up with an inconsistent notion of supervenience. To see this, consider the following argument: Assume (i)* and (ii)** are both satisfied. Now there are two possible cases: (1) For every P_k-property there is an A_i-property such that $Pr(A_i|P_k) = 1$, or (2) it is not true that (1). If (1) is the case, then (ii)** reduces to (ii)* and we end up with the original problem described in section 3. However, if (2) is the case, then there is a P_k-property that does not fully determine one specific A_i-property, meaning that there are at least two A_i-properties that could be realised by this P_k-property. This means, however, that the instantiation of P_k gets a probability greater than 0 conditional on the instantiation of each of these A_i-properties, which directly contradicts (i)*. Hence, we need to modify (i)* as well and, thus, arrive at the following relaxed version of probabilistic supervenience:

(SUP)** Agential properties $A_1, \ldots, A_m$ supervene on physical properties $P_1, \ldots, P_n$ if and only if

(i)** for every pair of non-identical properties A_i, A_j and every P_k: If $Pr(P_k|A_i) > 0$, then $Pr(P_k|A_i) \neq Pr(P_k|A_j)$, and

(ii)** for every P_k-property there is an A_i-property such that $Pr(A_i|P_k) \neq Pr(A_i)$.

The new condition (i)** relaxes the original first condition. In particular, it allows for more than one agential property to be instantiated by one and the same physical property, while still requiring that changing between instantiations of agential properties has (probabilistic) consequences in terms of changes in physical properties.

Let us further illustrate how the new version of supervenience **(SUP)**** works and how it can avoid the consequence of the collapse argument by having a look at Figure 3: The physical plane is identical to the physical plane in Figure 1, but the changed notion of supervenience now

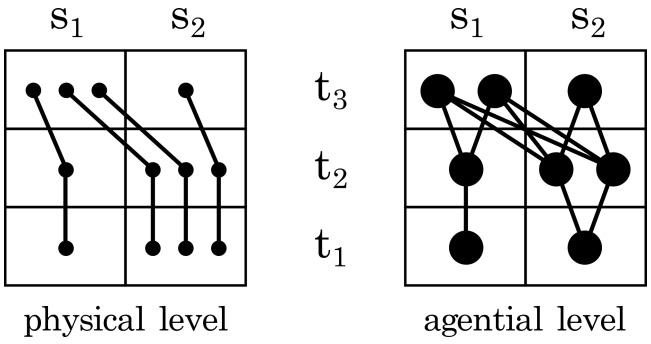

Figure 3

allows for a different picture of the agential plane. For our purposes, we assume that the physical properties instantiated in s_1t_1, s_1t_2, s_2t_1, and s_2t_3 give rise to the instantiation of the same agential properties as in Figure 1. The physical states in s_1t_3 and s_2t_2, however, can give rise to two different agential states. We assume that changing the agential state in any row changes the probabilities of all the physical states that could have realised that particular agential state, so (i)** is satisfied. Changing the agential state indicated by the left dot in s_2t_2 to the agential state indicated by the right dot in s_2t_2, for example, results in different probabilities for each of the physical states in s_2t_1 when conditioning on the different corresponding agential states in s_2t_2. When changing the agential state from one of those in s_2t_2 to the one in s_1t_2, on the other hand, the physical realisers of the initial state in s_2t_2 all get probability 0 and the physical state in s_1t_2 gets probability 1. We assume that for each possible physical state for which there is only one compatible agential state (i.e., the ones in s_1t_1, s_1t_2, s_2t_1, and s_2t_3), the probability of the agential state conditional on the corresponding physical state is 1. For the remaining cells in which there are more than one compatible physical state (i.e., s_1t_3 and s_2t_2), we assume that conditioning on each physical state gives rise to a probability of one of the corresponding agential states that is smaller than 1. Thus, also condition (ii)** is satisfied by the simple model in Figure 3.

Finally, we are in a position to revive compatibilist libertarianism. To see how, let us take another look at Figure 3. Like before, ontological non-identity and non-reducibility **(NR1)**, supervenience **(NR2)**, multiple realisability **(MR)**, and physical determinism **(DT)** are satisfied

by the model. This time, however, **(NR2)** relies on the probabilistic version of supervenience **(SUP)**** instead of the original version **(SUP)** (or its counterpart **(SUP)*** in terms of probabilities). Assuming, in accordance with **(SUP)****, that the physical states in s_1t_3 and s_2t_2 can each give rise to two different agential states, alters the possible world histories at the agential level. Here, the agential state in s_1t_2 can evolve into each one of the two different agential states in s_1t_3, the agential state in s_2t_1 can evolve into the two different agential states in s_2t_2, which can, in turn, each evolve into three possible states: the two in s_1t_3 and the one in s_2t_3. Now the problem with the initial version of compatibilist libertarianism was that once the physical state is taken into account, then all the possible future paths of action seemingly available at the agential plane collapsed into one. Hence, there was only one future path of action available to the agent, which meant, according to **(FW)**, that the agent did not have free will. Our revised version of compatibilist libertarianism can avoid this consequence. Even if the fact that the physical state at t_1 is fixed is taken into account, the future paths of action available to the agent at the agential plane do not necessarily collapse. If the physical state at t_1 is, for example, the one in s_1, then the two agential states represented by the dots in s_1t_3 are both accessible for the agent. The same goes if the physical state is the one represented by the left or the middle dot in s_2t_1. Only if the physical state is the one represented by the right dot in s_2t_1, is the agent determined to end up in the one single state in s_2 at t_3. However, even then there is a bit of freedom in the sense of **(FW)**: At s_2t_2, there are two agential states which are still accessible for the agent. All four possibilities are illustrated in Figure 4. Again, circles mark actual current states, black dots and lines stand for the states and trajectories accessible to the agent, and grey dots and lines stand for states and trajectories that are inaccessible to the agent.

Compatibilist libertarianism equipped with an interpretation of supervenience along the lines of **(SUP)**** finally achieves what the position's ultimate goal was: to block the incompatibilist argument introduced in section 2 by rendering P3 false. Our proposal succeeds where the original version of compatibilist libertarianism failed precisely because, unlike the original compatibilist libertarian account, it does not conflate branching of agential trajectories

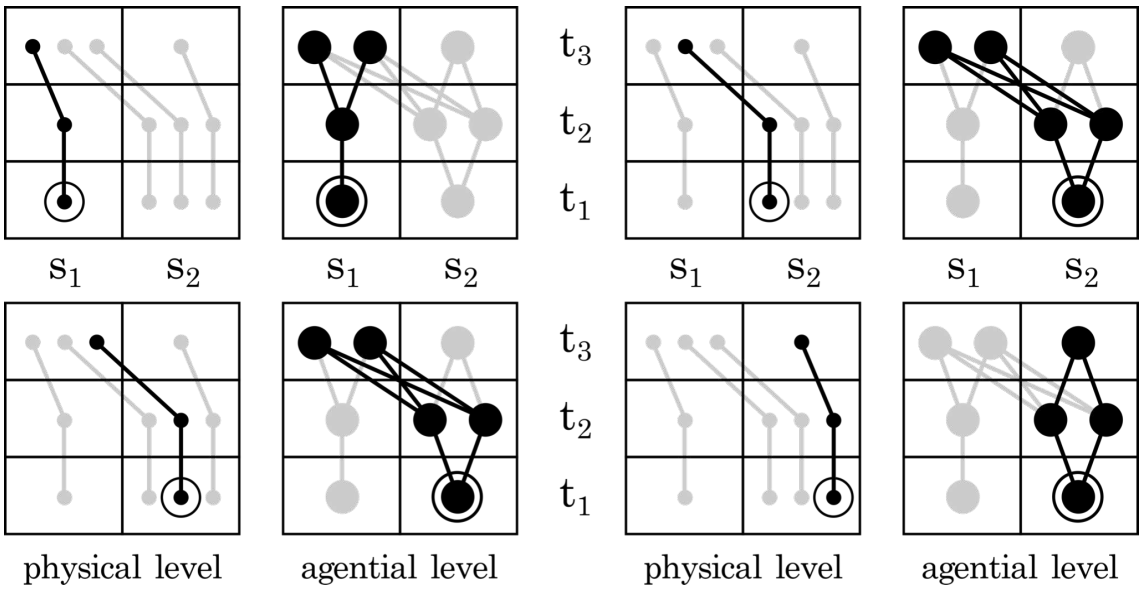

Figure 4

with courses of action actually available to the agent. According to our tweaked version of compatibilist libertarianism, branching at the agential level sometimes (but not necessarily always) opens up different courses of action that are actually accessible to the agent. The agential branches that are accessible are exactly those that are compatible with the fully determined future development of the actual physical state of the world (black agential trajectories in Figure 4). Agential branches that do not track the world's actual physical future (grey agential trajectories in Figure 4), on the other hand, are inaccessible to the agent. Note that this only opens up the *possibility* of actualist free will, while it does not yet establish that there *is* actualist free will.

5 Possible problems and responses

In this section we discuss possible objections to our modified version of compatibilist libertarianism and outline how a supporter of compatibilist libertarianism sympathetic to the weakened supervenience assumption **(SUP)**** could respond to them.

5.1 The agential states make no difference objection

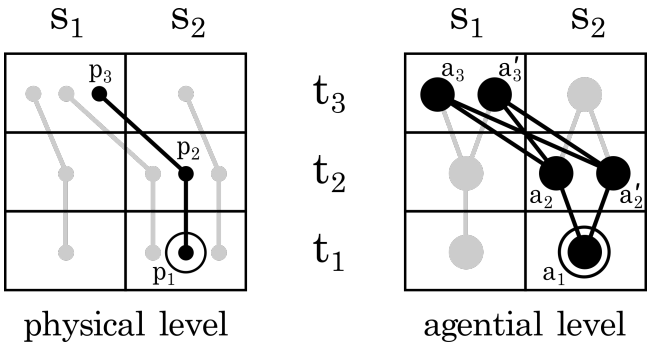

Figure 5

One possible objection to our new version of compatibilist libertarianism might go as follows: Agreed, branching at the agential level is now possible even if the actual physical state of the system at t_1 is taken into consideration. But look, this kind of freedom is still not worth much, and here is why: Ultimately, we want an actualist account of free will because we want our deliberation and voluntary decisions to be able to make a difference for how the future unfolds. But in the new version of compatibilist libertarianism, firstly, there is still no way for the agent to influence what is going on at the physical level. In some sense this was to be expected from the beginning when assuming physical determinism. But secondly, and much worse, the agent's agential state still contributes nothing to how the future unfolds. In the end, all the probabilities for the agent ending up in particular agential states are fully determined by the system's specific physical world history. If the world is, for instance, currently in the physical state p_1 represented by the dot in the circle in Figure 5, then it will ultimately evolve into the physical state p_3 at t_3 and the probabilities for the two compatible agential states in a_3 and a_3' at t_3 are fully determined by that specific physical state. Even in terms of probability it makes no difference whether the agent was in the agential state a_2 or a_2' at t_2. So our choices at the agential level still do not matter, neither for the system's future at the physical nor at the agential level. What we have now is still nothing but an illusion of free will.

We can defend the modified account against a charge like this as follows: We agree that one cannot influence what is going on at the physical level. This is simply a consequence of assuming physical determinism and there is no chance to escape it. But this was never what compatibilist libertarianism was aiming at. So what about the agential level? It does not follow from the

assumptions made so far that our particular agential state does not make any difference for how the future unfolds at the agential level. None of the assumptions made so far excludes that an agential state makes a probabilistic difference for a later agential state in addition to the physical state realising it. Let us have a brief look at the example given in the possible objection above: According to our assumptions, there will be specific conditional probabilities $Pr(a_3 \mid p_3)$ and $Pr(a_3' \mid p_3)$. However, nothing we assumed so far rules that the following two equalities hold:

$$Pr(a_3 \mid a_2, p_3) = Pr(a_3 \mid a_2', p_3) = Pr(a_3 \mid p_3)$$

$$Pr(a_3' \mid a_2, p_3) = Pr(a_3' \mid a_2', p_3) = Pr(a_3' \mid p_3)$$

These equalities would only follow from strict supervenience **(SUP)***. Hence, it is not only possible to have accessible branches at the agential level in the new version of compatibilist libertarianism, it is also possible that an agent's earlier agential states make a (probabilistic) difference for the later agential states that are possible given the particular physical state at the moment. And since the earlier physical state does not necessarily fully determine the earlier agential state anymore, the agent is at least in principle free to make up her mind and choose among the possible agential states compatible with the actual physical state. As mentioned before, this is not to establish that there *is* free will in the sense of compatibilist libertarianism, but rather to show that this type of free will is *possible* if one is ready to accept **(SUP**)**.

5.2 The anything goes objection

Another possible objection might go like this: Weakening supervenience as proposed in section 4 is overkill. It fully decouples what is going on at the agential level from what is going on at the physical level. Suddenly, any agential state becomes compatible with any physical state. This is way too much leeway for agential possibilities. Subscribing to **(SUP)**** would render the agential plane kind of freely floating above the physical plane. The original strict supervenience assumption **(SUP)** made sense precisely because it avoids this consequence. **(SUP)** properly

anchors the agential level in what is actually going on at the physical level; anything weaker than **(SUP)** cannot do the job.

There are several things one could say in response to such a line of reasoning.[12] First and foremost, subscribing to **(SUP)**** does not mean that anything is possible at the agential level. Like **(SUP)**, the main purpose of **(SUP)**** is to establish a minimal connection between the agential and the physical level. Both assumptions anchor the agential level by restricting what is agentially possible. While **(SUP)** implies that only one agential state is compatible with the actual physical state, **(SUP)**** in principle allows for more than one compatible state. However, **(SUP)**** does not allow for any agential state, by far. This can easily be seen by having a brief look at Figure 4. In each of the possible scenarios depicted, only some future agential trajectories are accessible to the agent, while others are inaccessible. The accessible ones are precisely the ones realisable by the already determined physical future. The resulting position is clearly a physicalist account committed to entity monism and property dualism. All the entities are still physical. Like in any non-reductive account, some of their properties are physical while others can be mental or agential. And, as said above, the higher-level states an agent can be in are, like in any other non-reductive account, clearly restrained by what is going on at the physical level. How strict the supervenience relation between the physical and the agential properties is is a question to be answered by nature and not a priori.

To further soften the more traditional view that higher-level states cannot be less than fully determined by physical states, let us compare the move from **(SUP)** to **(SUP)**** with the history of the philosophical debate about causation. In the early days, causation was also understood as a strict and often even as a necessitation relation. Causation was thought to be the metaphysical glue that holds reality together: The cause fully determines or even necessitates the effect. Any kind of relation weaker would not be able to do the job and would surely not deserve to be called causation. Over the years, however, it became obvious that many causes do not operate in this strict sense. Nowadays, causation is often understood as a kind of probabilistic difference-making relation (see, e.g., Cartwright, 1979; Pearl 2000; Schurz & Gebharter, 2016; Spirtes,

[12] For an argument why replacing **(SUP)** by **(SUP**)** does not undermine physicalism, see (Gebharter & Sekatskaya, 2024, sec. 5.3).

Glymour, and Scheines, 1993; Suppes, 1970) and almost no one seems to be too worried about that. Maybe it is time for a similar development with regards to the supervenience relation. A probabilistic concept of causation does a somewhat similar job as **(SUP)****: Causation links events together. The presence of the cause restricts the possible events resulting from it, it redistributes the probabilities of all the possible events following the cause. Similarly, **(SUP)**** links reality together, but vertically rather than horizontally. And similarly, the instantiation of the lower-level property restricts what is possible at higher levels by redistributing the probabilities of all the possible higher-level states.

5.3 The not enough goes objection

Another possible objection goes in exactly the opposite direction as the anything goes objection. According to it, the proposed version of compatibilist libertarianism still does not provide what actualists desire. Granted, one and the same physical state can now give rise to more than one agential state. But since the compatible agential states would share physical realisers, would they not be too similar to each other to be interesting? This is possible, but not necessary. We can suppose that sometimes compatible agential states would indeed be very similar, and perhaps would differ only in some insignificant details of their phenomenology. But **(SUP)**** allows for more significant and morally relevant differences as well. Imagine that Linda has just been offended by her friend Tina, and she is struggling with her feelings towards her friend. Should she forgive Tina, or should she hold a grudge?[13] Assume that all the other agential states are excluded by her actual brain state (which is fully determined by the actual physical state). Let the agential state a_1 in Figure 5 represent the psychological state of Linda's inner struggle, the agential state a_2 represent Linda forgiving her friend, and a_2' represent Linda holding a grudge. The weakened supervenience assumption **(SUP)**** allows that the same physical state gives rise to two different agential states, but with different probabilities. As we saw earlier, a

[13] Note that we are exclusively speaking about Linda's phenomenology here. Whether she forgives her friend or holds a grudge has the same physical consequences in the actual world due to physical determinism which is an integral part of the compatibilist libertarian strategy to block the incompatibilist argument (see section 2).

proponent of the modified version of compatibilist libertarianism can claim that earlier agential states have some probabilistic influence on which one of the different compatible later agential states will be realised. Consequently, whether a_2 or a_2' is realised at t_2 probabilistically depends not only on the subvenient physical state p_2, but also on what took place on the agential plane at t_1. This gives the agent the possibility to influence her future states. The fact that either a_2 or a_2' is realised at t_2 is probabilistically influenced by the agential state a_1. In our example, this means that whether or not Linda forgives Tina at t_2 probabilistically depends on Linda's inner struggle at t_1.

This modification of compatibilist libertarianism resembles Kane's (1996) event-causal libertarian account. According to Kane, agents can act quasi-deterministically and still possess free will. This happens when their acts are caused by the combined influence of their motives, their character, and the situation in which they find themselves (ibid., pp. 39f). However, in order to have free will and be morally responsible, sometimes an agent's acts must be truly undetermined. When both alternatives are open to the agent (in an actualist sense of open alternatives), they are both attractive for the agent for different, and incompatible motives, and the act of choosing one of these alternatives is the self-forming action (Kane, 1996, 2011, 2019). By choosing one alternative over the other, the agent makes and remakes herself into a particular kind of person. In other words, at the moments of self-forming actions, the agent not only chooses one of two or more alternatives, but also chooses what kind of character she develops over time. In such a way, the agent forms herself, for example, into a person who tends to be empathic or into a selfish person. There is a significant difference between Kane's account, which presupposes that the world is indeterministic at the physical level, and the compatibilist libertarian account, which presupposes only physical determinism. However, Kane's account of self-forming actions fits the assumptions of the modified version of compatibilist libertarianism. The only difference would be that for a compatibilist libertarian, the actualist alternatives take place at the agential level and the agent's decision does not influence the physical level, but the agential level. We can suppose that these acts of self-formation have a long-lasting influence on the probabilities of the outcomes of future agential choices.

Let us return to our example. If at t_2, Linda is in the agential state a_2 (i.e., the state of forgiving Tina), then the probability of Linda being calm at t_3 (represented by a_3) is higher than the probability of Linda being angry at t_3 (represented by a_3'). It might be objected that Linda does not have complete control over her future states. It might still happen that she is angry at t_3 even if she decided to forgive Tina at t_2. This, however, is not a problem for our account. It is enough that being calm is more likely after the decision to forgive than the other way round. Indeed, our own psychology seems to tell us that we do not have full control over our emotions. But it seems that we can exercise and develop self-control over time. And this is what our modified version of compatibilist libertarianism provides. It shows that the agential history as a whole might exercise probabilistic influence on the outcome of each future choice that the agent makes.

Moreover, even if it should ultimately turn out that we do not have significantly diverse agential states accessible at any given moment, or even if the influence of the prior agential states on posterior agential states is weak, the accumulation of such small choices and their weak influences might lead to large differences at the agential level in the course of the agential history. Think of a larger piece of the agential plane as the simple cases we depicted in our figures throughout this paper, displaying many years of an agent's lifespan and featuring thousands and thousands of branches, each one representing the rather similar consequences of a choice between two alternatives that look sufficiently similar. It would not be too surprising if some of these small choices or combinations of small choices lead to rather significant differences in outcome in the long run.

Finally, it can be objected that our version of compatibilist libertarianism gives very little room to agents because it restricts agential possibilities to the cases where these possibilities are compatible with the same physical state and such cases might not happen as often as our intuitive conception of free will demands. However, how widespread and diverse these agential possibilities are cannot be decided a priori. In the end, this is an empirical matter and we cannot simply assume that these cases occur rarely and that only the smallest changes would fit the modified compatibilist libertarian account. We admit that also the opposite cannot be claimed a priori. Anyway, what we suggest is, to the best of our knowledge, the only way one can

consistently hold an actualist position regarding free will in a world governed by determinism at the physical level. Thus, this seems the best actualist option we have right now and one should not discard it because it might turn out in the future that we have less freedom than we were hoping for. Another merit of the account is that it provides motivation to explore our actual degree of freedom. If one accepts it, then it makes sense to empirically investigate whether and how diverse agential states are compatible with identical physical states.

6 Conclusion

Compatibilist libertarianism promises to provide agents with actualist free will in a physically deterministic world. We have shown that while the original version of compatibilist libertarianism cannot live up to this promise, a modified version can do the job. The modification we have developed in this paper is achieved by weakening supervenience, one of the core assumptions of the compatibilist libertarian account. If supervenience is understood probabilistically, then agents can have alternative possibilities even if the future is completely determined at the physical level. Moreover, agents can in principle have sufficient control over their future by influencing the probability of the occurrence or instantiation of different agential states realised by the same physical states. We conclude that probabilistic supervenience allows compatibility of actualist free will with physical determinism.

Finally, we would like to stress that the results provided should be relevant and interesting even for readers who doubt that agential properties probabilistically supervene on physical properties in the actual world. Compatibilist libertarianism was proposed to explain how an agent can have actualist free will in a physically deterministic world. But as we showed, the original account cannot deliver what it promised. What needs to be done in order to get the desired compatibility of actualist free will and physical determinism is to replace strict supervenience by probabilistic supervenience. Now assume that one indeed has a good reason or even empirical evidence to believe that probabilistic supervenience does not hold in the actual world. Then, according to our analysis, one also has a good reason to reject an actualist account of free will along the lines of compatibilist libertarianism. According to our analysis in

section 4, there is not much one can do other than weakening supervenience in order to save compatibilist libertarianism from the collapse argument.

Acknowledgements: We would like to thank the participants of the workshop *Compatibilist libertarianism: Advantages and challenges* conducted at the Munich Center for Mathematical Philosophy in October 2021 for discussing the possibility of replacing strict by probabilistic supervenience with us and two anonymous reviewers.

References

Balaguer, M. (2010). *Free Will as an Open Scientific Problem*. MIT Press.

Baumgartner, M. (2010). Interventionism and epiphenomenalism. *Canadian Journal of Philosophy, 40*(3), 359–383.

Baumgartner, M. (2018). The inherent empirical underdetermination of mental causation. *Australasian Journal of Philosophy, 96*(2), 335–350.

Blanchard, T., & Hüttemann, A. (2024). Causal modeling in multilevel settings: A new proposal. *Philosophy and Phenomenological Research, 109*, 433–457.

Campbell, J. (1997). "A Compatibilist Theory of Alternative Possibilities." *Philosophical Studies* 88(3): 319–30.

Campbell, J. (2005). "Compatibilist Alternatives." *Canadian Journal of Philosophy* 35(3): 387–406.

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs, 13*(4), 419–437.

Dennett, D. C. (2003). *Freedom evolves*. Viking.

Ekstrom, L. (2000). *Free will: A philosophical study*. Focus Series. Westview Press.

Elzein, N., and T. Pernu. (2017). Supervenient freedom and the free will deadlock. *Disputatio* 9(45): 219–43.

Fara, M. (2008). Masked abilities and compatibilism. *Mind* 117 (468): 843–65.

Gebharter, A. (2017). Causal exclusion and causal Bayes nets. *Philosophy and Phenomenological Research, 95*(2), 353–375.

Gebharter, A. (2020). Free will as a higher-level phenomenon? *Thought: A Journal of Philosophy, 9*(3), 177–187.

Gebharter, A., & Sekatskaya, M. (2024). Mental causation, interventionism, and probabilistic supervenience. *Synthese, 203*, 206.

Ginet, C. (1990). *On Action*. Cambridge University Press.

Hausmann, M. (2020). The consequence of the consequence argument. *Kriterion – Journal of Philosophy, 34*(4), 45–70.

Hoffmann-Kolss, V. (2021). Interventionism and non-causal dependence relations: New work for a theory of supervenience. *Australasian Journal of Philosophy*. doi:10.1080/00048402.2021.1941154

Kane, R. (1996). *The significance of free will*. Oxford University Press.

Kane, R. (2011). Rethinking free will: New perspectives on an ancient problem. In *The Oxford Handbook of Free Will* (pp. 381–404). Oxford University Press.

Kane, R. (2019). The complex tapestry of free will: Striving will, indeterminism and volitional streams. *Synthese*, *196*(1), 145–160.

Kapitan, T. (2002). A master argument for incompatibilism? In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 127–157). Oxford University Press.

Kenny, A. (1978). *Freewill and responsibility*. Routledge & Kegan Paul.

Kim, J. (2005). *Physicalism, or something near enough*. Princeton University Press.

Lehrer, K. (1990. "A Possible Worlds Analysis of Freedom." In *Metamind*, 50–78. Oxford University Press.

List, C. (2014). Free will, determinism, and the possibility of doing otherwise. *Noûs, 48*(1), 156–178.

List, C. (2019a). What's wrong with the consequence argument: A compatibilist libertarian response. *Proceedings of the Aristotelian Society, 119*(3), 253–74.

List, C. (2019b). *Why free will is real.* Harvard University Press.

Menges, L. (2021). Free will, determinism, and the right levels of description. *Philosophical Explorations.* doi:10.1080/13869795.2021.1937679

Pearl, J. (2000). *Causality*. Cambridge University Press.

Pereboom, D. (2001). *Living without free will*. Cambridge Studies in Philosophy. Cambridge University Press.

Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford University Press.

Schurz, G., & Gebharter, A. (2016). Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets. *Synthese, 193*(4), 1073–1103.

Sekatskaya, M, & Gebharter, A. (2024). Supervenient fixity and agential possibilities. *Theoria.* doi:10.1111/theo.12551

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. Springer.

Suppes, P. (1970). *A probabilistic theory of causality*. North Holland.

Taylor, C., & Dennett, D. (2002). Who's afraid of determinism? Rethinking causes and possibilities. In R. Kane (Ed.), *Free will* (pp. 257–277). Blackwell Publishing.

van Inwagen, P. (1975). The incompatibility of free will and determinism. *Philosophical Studies, 27*, 185–199.

van Inwagen, P. (1983). *An essay on free will*. Oxford University Press.

Vihvelin, K. (2004). "Free Will Demystified: A Dispositional Account." *Philosophical Topics* 32(1): 427–50.

Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oxford University Press.

Woodward, J. F. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research, 91*(2), 303–347.

Woodward, J. (2022). Modeling interventions in multi-level causal systems: Supervenience, exclusion, and underdetermination. *European Journal for Philosophy of Science, 12*(4), 1–34.